

# La qualité dans les enquêtes

Anne-Marie Dussaix  
ESSEC Business School  
BP 50105 Cergy  
95021 Cergy-Pontoise Cedex  
e-mail : dussaix@essec.fr

## Résumé

Il existe une littérature académique très abondante sur la qualité des enquêtes. Mais il existe peu d'ouvrages traitant de l'ensemble des sources d'erreur dans une enquête. Les manuels de Théorie des Sondages sont essentiellement consacrés aux méthodes de sondage, c'est-à-dire aux méthodes de tirage de l'échantillon, aux méthodes d'estimation qui en découlent, et à l'erreur d'échantillonnage. Ils ne consacrent qu'un ou deux chapitres aux autres sources d'erreurs dans les enquêtes.

Cet article a pour objectifs de donner une vue d'ensemble des différentes sources d'erreur pouvant affecter la qualité des enquêtes et des méthodes possibles pour les prévenir et/ou y remédier, et de permettre ainsi aux utilisateurs de données d'enquête d'en évaluer la qualité.

## Abstract

This document deals with survey measurement quality, with particular emphasis on sampling and nonsampling errors. It tries to give to the users of survey data a clear understanding of how the quality of survey results can be affected by sampling and nonsampling errors and how to prevent and/or reduce these errors.

## 1. Introduction

La qualité des résultats obtenus dans une enquête par sondage dépend des choix effectués à chacune des étapes de réalisation de l'enquête, et pas seulement du nombre de réponses obtenues à l'enquête, même si ce dernier critère est souvent utilisé comme critère unique de qualité, que ce soit par ignorance ou par manque de connaissance sur la méthodologie de l'enquête.

Etant donnée l'importance des enquêtes par sondage dans de nombreux domaines, il est indispensable de mettre en œuvre des méthodes permettant d'assurer la qualité des résultats, de la contrôler et de l'évaluer.

Les cours classiques de Théorie des Sondages ou de Techniques de Sondage sont généralement consacrés aux méthodes de sondage, c'est-à-dire aux méthodes permettant de sélectionner un échantillon à partir d'une population finie, aux méthodes d'amélioration des estimations et aux méthodes de calcul des variances d'échantillonnage (Ardilly (2006), Tillé (2001)). Ces méthodes sont généralement exposées dans le cadre théorique d'un modèle où chaque individu sélectionné dans l'échantillon peut être joint et accepte de répondre à l'enquête (absence d'erreur de non-réponse), puis sait répondre, accepte de répondre aux questions posées et donne la vraie réponse (absence d'erreur de mesure). Dans ce contexte, le seul type d'erreur qui affecte les

résultats est *l'erreur d'échantillonnage*, c'est-à-dire la différence entre la valeur du paramètre étudié sur la population tout entière et la valeur du paramètre estimé sur l'échantillon sélectionné.

Dans la pratique des enquêtes par sondage, d'autres types d'erreur peuvent affecter les résultats. Les sources d'erreurs possibles sont par exemple :

- *les refus de participer à l'enquête*
- *le répondant* qui ne donne pas une réponse exacte à la question posée
- *l'enquêteur* qui ne respecte pas les consignes de passation du questionnaire, qui donne une définition erronée des termes employés dans le questionnaire, etc..
- *la saisie des données* si des erreurs sont faites lors de cette saisie.

Pour concevoir une enquête, évaluer et améliorer la qualité des résultats d'enquête, il est indispensable de connaître ces différentes sources d'erreur afin de savoir comment les prévenir et, *a posteriori*, si possible, comment en diminuer l'impact sur les résultats. Nous décrirons les sources d'erreur possibles et les moyens de les prévenir ou d'y remédier au §4. Nous présenterons préalablement les différentes étapes de réalisation d'une enquête au §2 et des définitions de la qualité au §3. Cette connaissance des différentes sources d'erreur permet également d'évaluer la qualité des résultats. Ceci fera l'objet du §5.

Cet article s'appuie largement sur les ouvrages de Dussaix et Grosbras (1996), Ardilly (2006) et Biemer et Lyberg (2006). Ce dernier ouvrage en particulier offre un exposé à la fois simple et complet sur la qualité dans les enquêtes.

## **2. Les différentes étapes dans une enquête par sondage**

Une enquête par sondage est une recherche d'informations, auprès d'une fraction de la population étudiée, appelée *échantillon*, le plus souvent au moyen d'un *questionnaire* ou parfois par observation directe (ex : panels de téléspectateurs, panels de distributeurs, ...), l'objectif étant d'estimer à partir des données collectées sur l'échantillon, certains paramètres caractéristiques de la population.

Il existe une très grande diversité des enquêtes par sondage. Nous nous placerons ici dans le contexte des *études quantitatives*, c'est-à-dire des études visant à quantifier les comportements, les opinions ou les attitudes d'une population à partir d'un échantillon. L'échantillon doit donc être d'une taille suffisante. Il doit être sélectionné de telle sorte qu'il permette l'extrapolation des résultats à la population tout entière, avec une précision acceptable, étant donnés les objectifs de l'enquête. La détermination de la taille d'échantillon en fonction de la précision souhaitée et les méthodes de sondage sont traités dans tous les manuels de Théorie des Sondages (cf. par exemple Ardilly, 2006) ; nous n'y reviendrons pas.

La qualité des résultats d'enquête dépend des choix qui ont été effectués à chacune des étapes de réalisation de l'enquête ; il est donc important de préciser ici ces différentes étapes.

On distingue généralement, dans une enquête par sondage avec collecte des informations réalisée par questionnaire, les étapes suivantes (cf. par exemple Dussaix et Grosbras, 1996) :

#### a) *Conception générale de l'enquête.*

Cette phase très importante comprend la définition des objectifs de l'enquête, de la population étudiée (ou population cible ou champ de l'enquête), des informations que l'on souhaite obtenir, avec quelle périodicité, à quel niveau de détail (pour quelles zones géographiques ? pour quelles périodes de temps ? sur quelles cibles ou sous-populations ?).

A ce stade, il est indispensable de s'assurer que les informations recherchées n'existent pas par ailleurs (données dites secondaires).

En fonction des points précédents, des coûts et des délais, en fonction aussi de l'existence ou non d'une base de sondage, de la longueur du questionnaire, on choisit une technique de collecte des informations (questionnaire ou observation). Ce choix a une influence directe sur la qualité des résultats.

Les différents modes d'administration du questionnaire sont principalement :

- par enquêteur (en face-à-face ou par téléphone)
- en auto-administré (essentiellement questionnaire envoyé par voie postale ou questionnaire par Internet dit « questionnaire en ligne »).

Le choix du mode de collecte est fondamental car il a une incidence sur le choix de la méthode de sondage et la forme du questionnaire. Ce choix dépend de différents critères. Les plus importants sont (Evrard et *al.* (2006), Jolibert et Jourdan (2006)) :

- le coût des différents modes de collecte (le face-à-face par enquêteur à domicile est le plus onéreux ; il s'impose cependant lorsque le questionnaire est long ou complexe),
- le questionnaire (sa longueur, sa complexité, la nécessité de montrer des supports visuels tels que des logos de magazine, l'importance des questions ouvertes),
- le taux de réponse anticipé,
- la dispersion géographique souhaitée : une enquête en face-à-face à domicile n'est pas le bon choix lorsque les objectifs de l'étude imposent une grande dispersion géographique des interviews.

Enfin, on détermine simultanément la méthode de sondage qui sera utilisée pour constituer l'échantillon. On détermine aussi la taille de l'échantillon nécessaire, compte tenu du budget disponible, de la précision souhaitée et des objectifs de l'enquête. Le choix de la méthode de sondage est un facteur déterminant dans la qualité d'une enquête. Un paragraphe sera consacré aux différentes méthodes de sondage et à l'impact du choix effectué sur la qualité.

#### b) *Rédaction du questionnaire.*

La rédaction proprement dite du questionnaire s'effectue généralement en trois étapes :

- première rédaction du questionnaire (si possible en utilisant les résultats d'études exploratoires préalables),
- pré-test du questionnaire auprès d'un petit nombre d'individus issus de la population cible afin de s'assurer de la de la bonne acceptation du questionnaire, de sa fluidité et de la bonne compréhension des questions,

- rédaction définitive incluant le pré-codage du questionnaire pour faciliter la saisie informatique.

### c) Administration du questionnaire

Le « terrain » est la phase de collecte des informations auprès des interviewés. Il commence par une réunion d'information des enquêteurs (ou briefing) afin de leur expliquer le sujet de l'étude, de leur transmettre les consignes d'échantillonnage et de passation du questionnaire ; il est suivi par l'administration des questionnaires sur le « terrain », et enfin par le contrôle de la qualité du travail des enquêteurs. Ce contrôle *a posteriori* est indispensable ; il permet de s'assurer de la matérialité des entretiens et du bon respect des consignes de recrutement des interviewés et de passation du questionnaire.

### d) Traitement et analyse des données

Cette dernière étape comprend :

- les opérations de vérification, de codage, de saisie informatique des questionnaires en cas de questionnaires « papier-crayon »,
- le traitement éventuel des questions ouvertes,
- le calcul des poids de sondage et des coefficients de redressement,
- le 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 ; les choix opérés à chacune des étapes ont une incidence directe sur la qualité globale de l'enquête.

## 3. Qu'est-ce que la qualité d'une enquête ?

Il n'existe pas de définition unique de la qualité des données d'enquête. Pour les statisticiens, la qualité est souvent assimilée à la précision des résultats calculée par la technique des intervalles de confiance. Mais, pour les utilisateurs des résultats, d'autres aspects sont importants comme, par exemple, la pertinence par rapport au problème posé, la mise à disposition des données en temps voulu.

Dans le manuel d'Eurostat sur l'évaluation de la qualité des données (*Handbook on Data Quality Assessment Methods and Tools*), trois aspects de la qualité des données sont considérés :

- les composantes de la qualité du produit statistique (au nombre de six : elles sont détaillées ci-dessous),
- la perception par les utilisateurs de la qualité des données,
- quelques caractéristiques du processus de production telles que la charge de travail demandée aux interviewés.

Dans le Système Statistique Européen, la qualité du produit statistique est évaluée selon les dimensions ou critères suivants (Eurostat, 2005) :

- la pertinence (*relevance*) est le degré auquel les résultats de l'enquête répondent aux besoins courants et potentiels des utilisateurs ;
- l'exactitude (*accuracy*) traduit la proximité des résultats ou des estimations avec la valeur vraie inconnue de la variable étudiée (l'information décrit bien le phénomène qu'elle doit mesurer) ; les erreurs d'échantillonnage et de non-échantillonnage qui seront décrites plus loin (*sampling and non-sampling errors*) affectent l'exactitude des résultats ;
- l'exactitude dans l'obtention des résultats et leur diffusion (*timeliness and punctuality*),
- l'accessibilité et l'intelligibilité (*accessibility and clarity*) des données statistiques,
- la comparabilité (*comparability*) est le degré auquel les différences de résultats entre zones géographiques ou entre périodes peuvent être attribuées à des différences réelles, et non pas à l'évolution des méthodologies ;
- la cohérence (*coherence*) est l'aptitude des données à être combinées de différentes façons et pour différentes utilisations.

Il existe d'autres versions un peu différentes de ces dimensions (Brackstone (1999) présente les critères utilisés à Statistique Canada).

Les composantes de la qualité qui sont privilégiées diffèrent selon la méthodologie d'enquête choisie. Une méthodologie d'enquête donnée ne peut pas optimiser toutes les composantes de la qualité simultanément ; par exemple, un compromis doit être trouvé entre la précision et les délais de réalisation de l'enquête.

Les dimensions de la qualité du produit statistique pourraient aussi être utilisées comme base pour l'évaluation de la qualité par les utilisateurs. Mais, dans de nombreux cas, les utilisateurs percevront la qualité de façon différente des producteurs (instituts de sondage, Instituts Nationaux de Statistique,...) et ne privilégieront pas les mêmes dimensions.

En ce qui concerne le dernier aspect de la qualité, c'est-à-dire la qualité du processus de production, il n'existe pas de définition standard. De façon très générale, la qualité du processus de production peut être évaluée à travers les points suivants :

- choix d'une méthodologie adaptée aux objectifs de l'étude
- réalisation du « terrain » selon les règles de l'art
- techniques d'analyse statistique appropriées (mise en œuvre correcte du calcul des poids de sondage et des coefficients de redressement, résultats calculés sur une taille d'échantillon suffisante,...).

En 1944, W.E. Deming, dans un article intitulé "On errors in surveys", citait treize facteurs pouvant affecter la validité des enquêtes ; le but recherché était d'insister sur la nécessité de diriger les efforts vers la réduction de ces treize sources d'erreur et non pas seulement vers une ou deux d'entr'elles. Il poursuivait en affirmant : «in the planning of a survey, effort should be directed toward the reduction of all the errors that it is possible to reduce, but the effort should be apportioned with a view to producing the greatest possible usefulness with the funds available », insistant ainsi sur la nécessité de relier les efforts à la qualité recherchée, et, de façon implicite, à l'utilisation qui sera faite des résultats.

Ceci reste un point majeur et difficile : il est souvent possible avec des coûts supplémentaires de réduire les biais dans une enquête mais la difficulté est bien dans l'ajustement adéquat des efforts déployés à l'utilisation qui sera faite des résultats.

Le sens traditionnel de qualité étant l'absence d'erreurs à chacune des étapes de réalisation de l'enquête (conception, plan de sondage, rédaction du questionnaire, terrain, analyses et conclusion), l'évaluation de la qualité d'une enquête passe par l'examen des sources d'erreur possibles à chacune des étapes de sa réalisation afin de les prévenir ou d'y remédier. C'est ce que nous détaillerons dans le paragraphe suivant.

#### 4. Les sources d'erreur dans une enquête par sondage

Une enquête par sondage permet d'estimer les valeurs inconnues de paramètres caractéristiques de la population (taux de notoriété, durée moyenne d'écoute,...). Ces valeurs estimées diffèrent des vraies valeurs inconnues à cause d'erreurs qui peuvent se produire à chacune des étapes de réalisation de l'enquête et créent *l'erreur totale* : l'erreur totale peut être définie simplement comme la différence entre la valeur estimée par l'enquête et la vraie valeur du paramètre de population.

Ces erreurs peuvent être réparties en deux groupes :

- l'erreur d'échantillonnage (*sampling error*),
- les autres types d'erreur (*nonsampling errors*) qui recouvrent essentiellement *l'erreur de mesure* et *l'erreur de non-réponse*.

Comment distinguer ces deux groupes d'erreurs ?

L'erreur d'échantillonnage est liée au processus d'échantillonnage et est intentionnelle : elle est évidemment maximale lorsqu'une seule unité est observée ; lorsque l'on observe un nombre croissant d'individus dans l'échantillon, elle décroît ; elle serait nulle si toute la population était observée (recensement). Pour la faire décroître, il suffirait d'avoir les moyens d'augmenter le nombre d'unités observées.

Remarquons que l'on parle classiquement d'erreur d'échantillonnage alors qu'il ne s'agit en aucun cas d'une erreur due à une mauvaise sélection de l'échantillon mais d'un aléa ou fluctuation due au principe même de l'échantillonnage.

A l'erreur d'échantillonnage, s'ajoutent d'autres types d'erreurs qui peuvent être considérés d'une certaine façon comme des erreurs non voulues et subies : les erreurs de mesure peuvent être dues par exemple aux interviewés qui, pour des raisons diverses, fournissent des réponses qui ne correspondent pas à la réalité de leurs opinions ou de leurs comportements ; elles peuvent être dues aussi aux enquêteurs qui peuvent donner aux interviewés une mauvaise interprétation des questions qui leur sont posées ou orienter les réponses. Ces erreurs ne diminuent généralement pas avec la taille d'échantillon ; elles peuvent même éventuellement augmenter avec la taille d'échantillon : il peut être par exemple plus difficile d'assurer un encadrement homogène du réseau d'enquêteurs lorsqu'un nombre important d'enquêteurs est nécessaire.

Comme nous l'avons dit précédemment, l'erreur d'échantillonnage est celle qui est le plus couramment considérée lorsque le sujet de la qualité d'une enquête est abordé. On sait quantifier

l'erreur d'échantillonnage ; il est par contre difficile de mesurer l'impact sur la qualité des résultats des autres types d'erreurs. Il n'y a pas de remède miracle contre elles, ce n'est qu'avec l'expérience qu'on apprend à les déceler, quelquefois les corriger ou les prévenir ; il faut pour cela comprendre leurs causes et connaître les moyens de les minimiser. Ce sera l'objet des paragraphes suivants.

Signalons qu'il existe une autre classification des erreurs dans une enquête par sondage en *erreurs de non-observation* (erreur de couverture, erreur d'échantillonnage, erreur de non-réponse) et *erreurs d'observation* (principalement erreur de mesure, erreurs de saisie et de codification).

#### **4.1. L'erreur d'échantillonnage**

Le calcul de cette erreur par la technique des intervalles de confiance fait l'objet des ouvrages de Théorie des Sondages.

Dans la mesure où les méthodes de sondage sont très largement décrites dans de nombreux manuels (par exemple Ardilly (2006), Dussaix et Grosbras (1996)), nous ne décrivons que très succinctement ces différentes méthodes en donnant des pistes pour le calcul des erreurs d'échantillonnage. Nous insisterons davantage sur l'importance du choix de la méthode pour la qualité globale de l'enquête.

Il existe en effet de très nombreuses façons de constituer un échantillon et, ce faisant, une grande diversité dans la façon dont les échantillons constitués permettent d'obtenir des résultats fiables. On assimile trop souvent « quantité » à « qualité », jugeant trop souvent la qualité d'un échantillon sur sa taille, c'est-à-dire sur le nombre de réponses obtenues alors qu'un « bon échantillon », sélectionné selon les règles de l'art, est de loin préférable à un « gros échantillon ».

On distingue très généralement deux grandes catégories de méthodes de sondage :

- *les méthodes aléatoires (ou probabilistes)* dans lesquelles chaque individu de la population étudiée a une probabilité connue, différente de zéro, donnée à l'avance, d'appartenir à l'échantillon ; cette probabilité est appelée probabilité d'inclusion ;
- *les méthodes empiriques (ou non aléatoires)* dans lesquelles les probabilités d'inclusion ne sont pas connues ; la constitution de l'échantillon résulte d'un choix raisonné visant à faire ressembler l'échantillon à la population dont il est issu.

Ces deux ensembles de méthodes seront décrites mais le calcul de l'erreur d'échantillonnage ne sera explicité que dans le seul cas du sondage aléatoire simple. Nous reportons le lecteur à des manuels de Théorie des Sondages pour le calcul de cette erreur d'échantillonnage dans le cas de sondages plus complexes.

Un troisième paragraphe exposera rapidement les techniques de redressement qui sont souvent utilisées *a posteriori* pour réduire l'erreur d'échantillonnage, mais aussi l'erreur de non-réponse et les erreurs de mesure.

##### **4.1.1. Les sondages aléatoires**

Ils peuvent être simples, stratifiés, à plusieurs degrés, à probabilités inégales.

Un sondage est **aléatoire** lorsque tout individu de la population étudiée a une probabilité connue, donnée à l'avance et différente de zéro d'appartenir à l'échantillon. Ces probabilités, appelées *probabilités d'inclusion*, ne sont pas nécessairement égales ; il suffit qu'elles soient définies *a priori*, que l'algorithme de tirage de l'échantillon respecte ces probabilités et que la méthode d'estimation en tienne compte.

Les sondages aléatoires ont trois *avantages* majeurs :

- la connaissance des probabilités d'inclusion de chaque individu de la population permet de calculer une marge d'erreur sur les résultats ;
- le tirage aléatoire fournit aux utilisateurs une protection contre une sélection biaisée de l'échantillon (mais les risques de biais dus au questionnaire ou aux non-réponses subsistent ; ces biais potentiels seront décrits plus loin) et rend les résultats « opposables aux tiers » ;
- la connaissance des taux de réponse fournit un critère d'appréciation de la qualité des résultats.

Les inconvénients des sondages aléatoires sont principalement les suivants :

- ils nécessitent de disposer d'une *base de sondage*, c'est-à-dire d'une liste des individus de la population ou au moins d'une liste de sous-populations ; ceci constitue le principal obstacle à l'utilisation des sondages aléatoires car il existe de nombreux cas où de telles listes n'existent pas. Par exemple, l'INSEE seule peut tirer des échantillons de foyers à partir de la base de sondage des logements en France ; cette base n'est pas disponible pour les instituts de sondage privés. Autre exemple : il n'existe pas de base de sondage des visiteurs d'un site touristique un jour donné ;
- ils sont généralement plus coûteux et demandent des délais de réalisation plus longs : une fois que l'échantillon a été sélectionné, il est en effet indispensable d'essayer d'obtenir une réponse des individus sélectionnés : par des revisites dans le cas d'absence dans les enquêtes à domicile, par des rappels à des heures et jours différents dans les enquêtes par téléphone, par des relances dans le cas d'enquête postales.

Décrivons rapidement les principales méthodes de sondage aléatoire, leurs avantages et inconvénients respectifs.

#### a) *Le sondage aléatoire simple*

Il consiste à tirer dans la population de taille  $N$  un échantillon de taille  $n$  sans remise de telle sorte que chaque individu ait la même probabilité d'inclusion, sans regroupement préalable ni manipulation sur la base de sondage ; il s'agit donc d'un simple tirage de numéros dans une urne. Ce tirage suppose que tous les individus de la population soient répertoriés sur une liste appelée *base de sondage*. Une façon simple de mettre en œuvre ce tirage est d'utiliser le tirage systématique : il consiste à tirer sur la base de sondage un individu sur  $k$  où  $k$  est le rapport taille de la population / taille de l'échantillon.

Le sondage aléatoire simple est en fait assez rarement utilisé :

- si la population est très hétérogène par rapport au problème étudié, les résultats obtenus par sondage aléatoire simple peuvent être peu précis. Si l'on dispose d'informations sur les

individus de la population permettant de les répartir en classes ou sous-populations plus homogènes, on aura intérêt à utiliser le sondage stratifié ;

- lorsque l'on ne dispose pas de la base de sondage et/ou pour réduire les coûts d'enquête, on utilise souvent le sondage en grappes ou le sondage à plusieurs degrés.

Le sondage aléatoire simple est cependant très important car il sert d'« ingrédient » de base dans des méthodes de sondage aléatoire plus complexes.

D'autre part, la précision des résultats dans un sondage aléatoire simple est facile à calculer et sert souvent d'approximation au calcul de la précision des résultats, notamment :

- lorsque d'autres méthodes plus complexes sont utilisées et que la présence d'erreurs de non-réponse et d'erreurs de mesure rendrait illusoire un calcul exact de précision,
- dans le cas de sondages empiriques où, théoriquement, la non-connaissance des probabilités d'inclusion ne permet pas de calculer la précision des résultats.

### ***Précision des résultats dans un sondage aléatoire simple***

Dans un sondage aléatoire simple, où l'objectif est d'estimer la valeur moyenne  $m$  d'une variable quantitative dans une population de taille  $N$ , si la taille d'échantillon est suffisamment grande ( $n > 30$ ), l'intervalle de confiance au degré de confiance 95% est estimé par :

$$(1) \quad \bar{y} - 1,96 \frac{s}{\sqrt{n}} \sqrt{1 - \frac{n}{N}} < m < \bar{y} + 1,96 \frac{s}{\sqrt{n}} \sqrt{1 - \frac{n}{N}}$$

où  $\bar{y}$  et  $s$  désignent respectivement la moyenne et l'écart-type de la variable étudiée dans l'échantillon.

Ce résultat indique donc que la précision du sondage (définie comme le demi-intervalle de confiance correspondant à un degré de confiance donné) dépend :

- de l'écart-type de la variable étudiée,
- de la taille de l'échantillon,
- de la taille de la population par l'intermédiaire du taux de sondage  $n/N$ .

Mais si ce taux de sondage est faible (en pratique, inférieur à 10%), on peut considérer que le facteur d'exhaustivité  $(1 - n/N)^{0,5}$  est négligeable (en effet, il est supérieur à 0,949 si  $n/N$  est inférieur à 10%). L'intervalle de confiance au degré de confiance 95% est alors :

$$(2) \quad \bar{y} - 1,96 \frac{s}{\sqrt{n}} < m < \bar{y} + 1,96 \frac{s}{\sqrt{n}}$$

Dans ce cas (taux de sondage inférieur à 10%), la précision du sondage ne dépend pas du taux de sondage. Elle ne dépend que de la taille de l'échantillon et de l'écart-type de la variable étudiée : un sondage portant sur 1 000 individus dans une population de 50 millions d'habitants donnera la même précision qu'un sondage de même taille dans une population de 100 000 habitants,

à méthode d'enquête et en supposant que la dispersion de la variable étudiée est identique dans les deux populations.

Les formules (1) et (2) permettent également de calculer *a priori* avant de réaliser le sondage la taille d'échantillon nécessaire pour estimer la moyenne  $m$  dans la population avec une précision inférieure à une précision  $D$  donnée à l'avance. La formule (2) permet de calculer cette taille d'échantillon, dans le cas où le facteur d'exhaustivité est négligeable :

$$n \geq \frac{(1,96)^2 s^2}{D^2}$$

Dans ce cas,  $s$  désigne une estimation de l'écart-type de la population effectuée avant la réalisation du sondage et obtenue par un sondage antérieur ou une pré-enquête.

Lorsque l'objectif du sondage est d'estimer une proportion  $p$  d'individus dans la population possédant un caractère donné, l'intervalle de confiance au degré de confiance 95% pour  $p$  s'écrit, lorsque le facteur d'exhaustivité est négligeable et pour  $n$  assez grand :

$$f - 1,96\sqrt{\frac{f(1-f)}{n}} < p < f + 1,96\sqrt{\frac{f(1-f)}{n}}$$

Dans cet intervalle de confiance,  $f$  est la proportion d'individus possédant le caractère donné dans l'échantillon.

A titre d'exemple, le tableau suivant donne le demi-intervalle de confiance au degré de confiance 95% correspondant à différentes tailles d'échantillon lorsque la proportion  $f$  dans l'échantillon prend les valeurs indiquées dans les colonnes du tableau.

Proportion f  Taille d'échantillon	5%	10%	15%	20%	30%	40%	50%
	ou 95%	ou 90%	ou 85%	ou 80%	ou 70%	ou 60%	
100				7,8	9,0	9,6	9,8
150			5,7	6,4	7,3	7,8	8,0
200		4,2	4,9	5,5	6,4	6,8	6,9
300	2,5	3,4	4,0	4,5	5,2	5,5	5,7
400	2,1	2,9	3,5	3,9	4,5	4,8	4,9
500	1,9	2,6	3,1	3,5	4,0	4,3	4,4
700	1,6	2,2	2,6	3,0	3,4	3,6	3,7
1000	1,4	1,9	2,2	2,5	2,8	3,0	3,1
2000	1,0	1,3	1,6	1,8	2,0	2,1	2,2
5000	0,6	0,8	1,0	1,1	1,3	1,4	1,4
10000	0,4	0,6	0,7	0,8	0,9	1,0	1,0

Rappelons que cette précision donnée par le demi-intervalle de confiance ne fait référence qu'à l'erreur d'échantillonnage et pas aux autres types d'erreur.

## b) Le sondage aléatoire stratifié

Dans le cas où les variables étudiées présentent une dispersion forte dans la population, la précision des résultats d'un sondage aléatoire simple peut être améliorée par utilisation du *sondage aléatoire stratifié*.

Le principe de la *stratification* est de découper la population en sous-ensembles appelés *strates* (ou d'utiliser une partition existante) et de réaliser un sondage dans chacune de ces strates.

Si les strates sont plus homogènes que la population par rapport aux variables étudiées, les estimations seront plus précises que celles qui seraient obtenues par sondage aléatoire simple pour une taille d'échantillon identique.

C'est ainsi que les échantillons d'individus ou de ménages sont stratifiés par région croisée par type d'habitat (taille des communes) ; les échantillons d'entreprises sont stratifiés par secteur et par taille, exprimée en effectifs salariés ou chiffre d'affaires ; les échantillons d'exploitations agricoles par tranches de surface, les échantillons de jeunes sortis de l'enseignement supérieur par discipline, etc. Pour l'étude du lancement d'un nouveau produit financier, il sera pertinent d'utiliser des critères liés aux revenus, à l'âge, éventuellement au sexe, c'est-à-dire à des facteurs susceptibles d'expliquer les différences de comportement financier.

La stratification permet donc de contrôler l'échantillon et d'améliorer ainsi la précision des estimations obtenues dans le sondage en éliminant la possibilité de tirer des échantillons dans lesquels certaines classes seraient sous- ou surreprésentées par le fait du tirage aléatoire. Dans le cas où la population est stratifiée par régions, par exemple, on pourra répartir l'échantillon global entre les différentes régions proportionnellement à la population de chaque région : c'est ce que l'on appelle un *échantillon stratifié proportionnel*. Mais cette répartition, pour naturelle qu'elle soit, n'est pas toujours la plus pertinente. Le choix de la répartition dépend essentiellement des objectifs de l'enquête :

- lorsque le sondage a pour objectif de fournir des estimations au niveau de la population globale et de certaines strates, on peut surreprésenter celles-ci dans l'échantillon (à condition d'en tenir compte dans les procédures d'estimation) ;
- lorsque le sondage a pour seul objectif de fournir au niveau de la population des estimations avec la meilleure précision possible, la meilleure répartition est *la répartition dite optimale ou répartition de Neyman* dans laquelle la taille d'échantillon dans une strate est proportionnelle au nombre d'individus dans la strate **et** à l'écart-type de la variable étudiée dans la strate (cf. Ardilly (2006) pour plus de détails).

Pour réaliser un sondage aléatoire simple ou un sondage aléatoire stratifié, il est nécessaire de disposer de la base de sondage. D'autre part, même si l'on dispose de cette base de sondage, le coût de l'enquête peut être prohibitif en raison des coûts de déplacement lorsque la population est dispersée géographiquement et que l'enquête est réalisée par enquêteur à domicile.

On peut alors avoir recours à deux autres méthodes de sondage aléatoire : le sondage en grappes et le sondage à plusieurs degrés.

### c) *Le sondage en grappes et le sondage à plusieurs degrés*

La pratique des sondages par grappes ou à plusieurs degrés est très largement répandue ; elle est motivée par la nature des données à recueillir, des considérations de coût ou de faisabilité, la mauvaise qualité, voire l'inexistence des bases de sondage.

En effet, dans de très nombreux exemples de sondages, la population peut être considérée comme répartie en groupes (appelées grappes ou Unités Primaires) : les ménages sont des grappes d'individus, les vols des compagnies aériennes sont des grappes de passagers, les immeubles sont des grappes de logements,....

On tire alors un échantillon aléatoire de « grappes » :

- si tous les individus des « grappes » tirées sont sélectionnés, le sondage est dit en grappes ;
- si l'on observe seulement un échantillon dans chacune des « grappes » tirées, on parle de sondage à degrés.

Ce type de sondage

- permet de résoudre le problème de manque de base de sondage (tout en réalisant un sondage aléatoire),
- permet généralement de réduire les coûts d'enquête.

Par contre, si les individus à l'intérieur d'une même grappe ou d'une même unité primaire ont tendance à se ressembler en ce qui concerne les thèmes de l'enquête, les résultats peuvent être moins précis que dans un sondage aléatoire simple de même taille (mais la réduction des coûts permet d'augmenter la taille d'échantillon).

Donnons quelques exemples :

- 1) *Etudes médicales.* Certaines études sont réalisées auprès d'échantillons de médecins qui sont ainsi considérés, pour l'enquête, comme des grappes de patients ou de prescriptions.
- 2) *Sondages électoraux.* Les soirs de consultations électorales, les sondages « sortie des urnes » sont réalisés auprès d'électeurs à la sortie de bureaux de vote. Il s'agit de sondages à deux degrés, le premier degré consistant à choisir les bureaux de vote où opéreront les enquêteurs.
- 3) *Enquêtes INSEE.* Les échantillons de ménages ou d'individus réalisés par l'INSEE sont des exemples de sondages à trois ou quatre degrés. En général, les unités finales de sondage sont les logements, et leur tirage est organisé comme suit :
  - on procède à un tirage de cantons ou d'agglomérations, avec des probabilités inégales proportionnelles à leur population, et stratification préalable par région et type d'habitat (zones rurales, tailles d'unités urbaines), c'est le premier degré de sondage;
  - à l'intérieur de ces unités primaires, on effectue un tirage de communes, ou de quartiers, unités secondaires de sondage, toujours à probabilités inégales proportionnelles à leur population;
  - enfin, on tire au sort des logements dans les unités secondaires, selon un sondage aléatoire simple, généralement le même nombre de logements par unité secondaire.

Lorsque l'objet du sondage a trait aux individus, comme les dépenses de loisirs ou de santé, le ménage est alors considéré comme une *grappe* d'individus, d'où un quatrième degré de sondage pour sélectionner l'individu à interroger.

#### 4.1.2. Les sondages non aléatoires ou empiriques

Les méthodes de sondage aléatoire qui ont été décrites précédemment supposent le tirage *aléatoire* de l'échantillon à partir d'une *base de sondage*, c'est-à-dire d'une liste exhaustive des individus composant la population étudiée. Lorsque de telles bases sont inexistantes ou indisponibles, ou lorsqu'il est trop coûteux de réaliser un sondage aléatoire, on a recours aux *méthodes dites non aléatoires*, ou encore méthodes *empiriques* ou à *choix raisonné*.

Un sondage est *non aléatoire* lorsqu'il n'inclut pas de mécanisme de sélection aléatoire des individus de la population : la probabilité de sélection des individus de la population n'est donc pas connue, peut même être nulle pour certains d'où l'existence de biais dits de sélection, et l'impossibilité de calculer des marges d'erreur.

Les méthodes non aléatoires, essentiellement la *méthode des quotas*, sont cependant très utilisées dans les sondages d'opinion et les études de marché pour deux raisons principales.

- la rareté ou non-disponibilité des bases de sondage

En France, les instituts de sondage privés ne disposent généralement pas de base de sondage pour tirer leurs échantillons : la base de sondage des logements issue du recensement est couverte par le secret statistique. On pourrait penser aux listes électorales (dans le cas de sondage auprès des Français depuis de 18 ans) ; mais, le vote n'étant pas obligatoire en France, elles sont trop inexactes. Reste l'annuaire des abonnés à France Télécom qui peut effectivement être utilisé comme base de sondage dans le cas d'enquêtes par téléphone, malgré ses imperfections croissantes (abonnés sur liste rouge, exclusifs portables,... ) qui seront décrites plus loin.

Autre exemple : les enquêtes réalisées en centre commercial, sur site touristique pour lesquelles on ne dispose pas de base de sondage.

- le coût et les délais de réalisation

Dans le cas d'enquête par enquêteur en face-à-face, on considère qu'une enquête aléatoire sur adresses coûte environ trois fois plus cher qu'une enquête par quotas (à nombre de points d'enquête tirés équivalent).

##### a) *La méthode des quotas*

Parmi les méthodes non aléatoires, la *méthode des quotas* est certainement la plus utilisée dans les sondages d'opinion et les études de marketing.

Le principe de la méthode des quotas est de construire une maquette, un modèle réduit de la population étudiée, selon des critères dont on connaît la répartition dans la population.

Le *principe* de la méthode est le suivant :

- on choisit quelques caractéristiques dont on connaît la distribution dans la population étudiée ; par exemple, dans les enquêtes sur individus ou sur ménages, ce seront généralement les critères sociodémographiques correspondant aux statistiques publiées après recensement par les instituts nationaux de statistique ;
- on donne à chaque enquêteur un plan de travail qui lui impose le respect de certaines proportions au sein de ses interviews.

La structure de la population étudiée selon ces critères est connue :

- à partir de statistiques disponibles (enquêtes de l'INSEE,...)
- à partir d'une enquête *ad hoc* de préférence de grande taille dite enquête de cadrage (par exemple, dans le panel d'internautes Médiamétrie/NetRatings, l'étude de cadrage « L'Observatoire des Usages Internet » (1 000 interviews par téléphone et par mois) permet d'estimer le nombre et le profil des individus ayant accès à Internet).

### *Exemple*

L'étude AEPM, réalisée pour le compte d'Audipresse, a pour objectif de mesurer et de qualifier l'audience de la presse magazine au niveau national sur la base d'un cumul de 12 mois d'enquête (168 magazines étudiés sur l'ensemble de l'année 2007).

L'univers de l'enquête (ou population cible) est constitué de l'ensemble des individus âgés de 15 ans ou plus, appartenant à un ménage ordinaire et résidant en France.

L'enquête est réalisée en face à face à domicile sur CAPI (*Computer Assisted Personal Interview*) selon la méthode des quotas.

Le terrain est confié à trois instituts : IPSOS, ISL et TNS Sofres, avec une répartition en trois sous-échantillons homogènes (dispersion géographique et contraintes méthodologiques identiques). Il a lieu en continu sur l'ensemble de l'année.

L'échantillon total est de 24 000 individus âgés de 15 ans ou plus dont :

- ⇒ un échantillon principal de 20 100 individus représentatif de la population de l'étude,
- ⇒ un sur-échantillon de 3 900 individus appartenant à un foyer dont le chef de famille est "cadre" ou "profession intermédiaire".

Le plan de sondage sera uniquement décrit pour l'échantillon principal :

- tirage aléatoire dans le fichier BDCOM de l'INSEE de points d'enquête à partir d'une matrice régions (21) x catégories d'habitat (9) pour l'ensemble des 6 vagues d'enquête.
- répartition des points d'enquête en 126 sous-ensembles (6 vagues x 3 instituts x 7 jours nommés).

Dans chacun des points d'enquête :

- dix interviews à réaliser sur deux jours, pour l'échantillon principal
- quatre interviews à réaliser sur un ou deux jours, pour le sur-échantillon

Dans les communes de 10 000 habitants ou plus, pour l'échantillon principal, tirage aléatoire d'îlots (six îlots par point d'enquête), afin d'assurer une bonne dispersion géographique des interviews à l'intérieur de la commune

Désignation de la personne à interroger selon la méthode des quotas (calculés par région INSEE sur la base du Recensement 99 et de l'Enquête Emploi 2002 pour la PCS).

LES VARIABLES DE QUOTAS pour l'échantillon principal sont les suivantes :

- Sexe x actif/inactif

- Age (7 classes : 15-17 ans / 18-21 ans / 22-34 ans / 35-49 ans / 50-64 ans / femmes âgées de 65 ans ou plus / hommes âgés de 65 ans ou plus)

- PCS de l'interviewé (6 ou 8)

Nombre limité de chômeurs autorisé (un ou deux par point d'enquête selon les départements)

Le choix des personnes interrogées est laissé à l'initiative de l'enquêteur sous réserve qu'il respecte la répartition fixée.

On notera que l'échantillon obtenu est ainsi représentatif de la population *par rapport* aux variables de quotas en ce sens qu'il respecte les proportions constatées dans la population sur ces variables. Rien n'indique cependant qu'il soit représentatif sur d'autres critères et sur les variables qui font l'objet de l'étude.

La fiabilité des résultats d'une enquête par quotas (à domicile en face à face) dépend beaucoup de la façon dont elle est *mise en œuvre* ; l'expression « sondage réalisé par la méthode des quotas » peut recouvrir des réalités de mise en œuvre très différentes. Cette fiabilité dépend en particulier du mode de sélection des communes dans lesquelles vont opérer les enquêteurs. Le tirage aléatoire de ces communes est une garantie contre une insuffisante dispersion géographique des interviews ; dans certaines enquêtes, on améliore encore cette dispersion par tirage aléatoire des îlots où doivent avoir lieu les interviews (c'est le cas de l'enquête AEPM). Cela peut être important dans les enquêtes où l'on s'intéresse à des comportements liés à la mobilité. Il est par ailleurs important de contrôler le travail des enquêteurs par contre-enquête (contre-visite, contrôle par lettre ou par téléphone).

Pratiquement, on limite la liberté de choix des enquêteurs en ajoutant à la feuille de *quotas* des *consignes de recherche* afin de reproduire le plus possible un tirage équiprobable :

- ne pas effectuer plus d'une interview par immeuble ou dans plus d'une maison individuelle par rue,
- varier les étages dans lesquels sont réalisées les interviews,
- changer d'îlot ou de quartier après chaque enquête,
- respecter la répartition indiquée des interviews dans la journée et en soirée,
- ne pas interviewer une personne interrogée depuis moins d'un an

Comme dans la stratification, les quotas doivent être liés au sujet étudié mais il ne faut pas les multiplier afin de ne pas rendre le travail de l'enquêteur tellement difficile en fin de quotas qu'il

serait conduit soit à ne pas remplir ses quotas (ce qui est acceptable dans une limite raisonnable), soit à tricher. Mieux vaut alors utiliser des variables jugées importantes comme variables de redressement, en pensant à les inclure dans le questionnaire.

Enfin, comme on l'a vu plus haut, on donne généralement à l'enquêteur non pas des quotas *croisés* mais des quotas *marginiaux* : avec des enquêteurs expérimentés, ceci diminue le temps de recherche des interviewés.

Les *inconvénients* de la méthode des quotas sont ceux des méthodes *non aléatoires* en général :

- existence de biais que ne peuvent éviter les consignes de recherche complétant généralement les quotas. En particulier, si le nombre de refus qu'a dû subir l'enquêteur est élevé, la représentativité peut être sujette à caution. La qualité des enquêtes par quotas repose essentiellement sur la qualité du travail de l'enquêteur.
- impossibilité de calculer des marges d'erreur ; on donne généralement comme approximation grossière, faute de mieux, les marges d'erreur du sondage aléatoire simple données au §4.1.1.

Les *avantages* de la méthode des quotas sont essentiellement :

- un coût et des délais de réalisation plus faibles que ceux d'une enquête aléatoire, ce qui justifie son emploi lorsqu'erreurs de mesure et erreurs de non-réponse sont difficilement compressibles comme dans les sondages d'opinion.

#### b) *Autres méthodes de sondage non aléatoire*

Ce sont principalement la méthode des unités-types, peu employée, qui ne sera pas décrite et l'échantillonnage de volontaires.

### **L'échantillonnage de volontaires**

Il s'agit d'enquêtes réalisées auprès de lecteurs de tel ou tel journal, d'adhérents d'associations acceptant de répondre à un questionnaire, de téléspectateurs acceptant de répondre par téléphone ou par Internet à une ou plusieurs questions posées au cours d'une émission. C'est le cas aussi des internautes qui répondent à un questionnaire proposé sur un site. L'extrapolation des résultats obtenus à la population étudiée doit être faite avec précaution, en essayant d'apprécier qualitativement l'impact des biais de sélection en fonction du sujet de l'étude. Par exemple, on sait que le recrutement sur Internet a tendance à surreprésenter les gros consommateurs d'Internet ; il s'agit alors de déterminer si ce biais peut avoir un impact sur les résultats, et, si oui, d'essayer de le corriger par des procédures de redressement.

On peut aussi classer dans cette catégorie les échantillons obtenus sur « Access Panels ». Les Access Panels sont des échantillons, souvent de très grande taille, formés d'individus pré-recrutés par un institut de sondage et acceptant le principe d'être interrogés s'ils sont sollicités. Ils peuvent être interrogés par voie postale, par téléphone ou sur Internet. Leur grande taille permet d'identifier un nombre suffisant de consommateurs d'un produit ou service à faible taux de pénétration ou acheté rarement. Le développement des Access Panels est dû aux difficultés

croissantes de recrutement d'échantillons *ad hoc* et à la multiplication des produits qui rendent nécessaires les études sur cible fine. Cependant, notamment dans le cas des Access panels sur Internet, la représentativité de l'échantillon des répondants doit être étudiée au cas par cas en fonction du thème de l'enquête et de la population étudiée.

#### 4.1.3. Les méthodes de redressement

Parmi les méthodes de sondage que nous avons décrites, la stratification permet d'améliorer la précision des résultats en utilisant des informations dont on dispose sur la population. Cette connaissance de la population permet de définir des modalités de tirage plus efficaces.

Les méthodes de redressement sont utilisées *a posteriori*, lorsque le sondage a été réalisé, pour réduire l'erreur d'échantillonnage et/ou l'erreur de non-réponse (on les appelle alors méthodes de repondération) et, éventuellement, les erreurs de mesure.

Le principe en est le suivant :

On dispose :

- de résultats obtenus sur un échantillon (simple, stratifié, à degrés,...) dans une enquête par sondage
- d'informations connues sur la population étudiée (par exemple : sexe et âge des médecins dans une enquête auprès de médecins).

On constate que la structure de l'échantillon diffère de celle de la population (à cause de fluctuations d'échantillonnage, de non-réponse à l'enquête qui affecte différemment différentes catégories de population, d'erreurs de mesure,...)

Les méthodes de redressement permettent d'améliorer les résultats en tenant compte des informations connues sur la population et des différences constatées. Elles consistent à affecter des poids aux réponses des interviewés dans l'analyse des résultats de telle sorte que l'on retrouve alors les proportions connues sur la population.

Notons que dans les enquêtes par sondage, si l'échantillon est tiré à probabilités inégales, on introduit aussi généralement des « poids de sondage » dans l'analyse des résultats (ces poids de sondage sont l'inverse des probabilités d'inclusion). Les poids issus du redressement vont modifier les poids de sondage avec un objectif décrit plus loin.

Plus précisément, les objectifs du redressement d'enquêtes sont :

- améliorer la précision des estimations (réduire les intervalles de confiance),
- faire en sorte que la structure de l'échantillon soit identique à la structure connue sur la population de certaines variables liées aux variables étudiées, donc limiter les contestations,
- corriger partiellement le biais dû aux non-réponses (ou, plus rarement, aux erreurs de mesure).

*Remarque* : si l'échantillon est « correct », le redressement modifie peu les résultats.

Il existe de nombreuses méthodes ; leur choix est lié à la nature des variables de redressement (qualitatives *versus* quantitatives) et à leur nombre (une *versus* plusieurs)

*Exemple : la stratification a posteriori ou post-stratification (redressement sur un seul critère qualitatif)*

On considère un échantillon de 1 000 individus de 15 ans et plus interrogés sur la lecture du magazine Rond-Point au cours des six mois précédant l'enquête.

Sur le critère « Rural-Urbain » pour lequel on a constaté dans des enquêtes antérieures qu'il était lié à la lecture du magazine, on sait que la proportion des individus habitant en zone rurale dans la *population* est 25%. Dans l'*échantillon*, cette proportion n'est que 22%.

La stratification *a posteriori* consiste à calculer les résultats :

- en donnant un poids (appelé coefficient de redressement) de  $0,25 / 0,22 = 1,136$  aux questionnaires « zone rurale »
- en donnant un poids de  $0,75 / 0,78 = 0,962$  aux questionnaires « zone urbaine ».

On vérifie que la proportion de ruraux est alors estimée par :

$$\sum_{i=1}^{220} 1,136 / 1000 \cong \sum_{i=1}^{220} \frac{0,25}{0,22 \times 1000} = 0,25$$

Lorsque l'on souhaite redresser l'échantillon sur plusieurs critères dont on ne connaît pas la structure croisée, des algorithmes (algorithme du *raking ratio* ou méthode de calage généralisée, cf. Ardilly (2006)) permettent de déterminer un *poids unique* par individu qui est alors utilisé dans le calcul des résultats de l'enquête

Le redressement sur critères multiples est utilisé dans toutes les « grosses » enquêtes.

## 4.2. Les autres types d'erreurs

Les autres types d'erreurs incluent principalement l'erreur de couverture, l'erreur de non-réponse et les erreurs de mesure.

*L'erreur de couverture* provient essentiellement de la différence entre la valeur du paramètre sur la *population cible* ou population étudiée, et la valeur du paramètre sur la base de sondage à partir de laquelle est sélectionné l'échantillon.

*L'erreur de non-réponse* provient de l'absence de réponse de la part d'individus appartenant à l'échantillon sélectionné.

*L'erreur de mesure* (ou d'observation) provient d'écarts entre les réponses fournies par les interviewés et les vraies valeurs.

Elle peut être due :

- aux enquêteurs (d'où formation et contrôle nécessaires),
- à l'instrument de mesure (mode d'enquête, questionnaire,...),
- à l'interviewé (il fournit une information incorrecte, volontairement ou involontairement).

D'autres types d'erreurs (non décrites dans cet article) peuvent se produire :

- erreurs de codage et de saisie
- erreurs liées à l'imprécision des informations utilisées pour la construction de l'échantillon ou lors des procédures de redressement (exemple : redressement d'un panel d'internautes sur une enquête de cadrage).

Les tendances observées sont :

- une *diminution des erreurs de mesure* à cause du développement important depuis deux décennies de nouvelles technologies de collecte d'information (collecte assistée par ordinateur, audimètres, lecteurs de codes à barre dans les panels de consommateurs et de distributeurs,...)
- une *augmentation des erreurs de réponse* dues à l'augmentation très largement observée des taux de non-réponse dans les enquêtes (cf. Brignier et Dupont, 2005).

#### 4.2.1. L'erreur de couverture

L'*erreur de couverture* est l'erreur causée par la différence entre :

- la valeur du paramètre sur *la population cible* ou population étudiée, et
- la valeur du paramètre sur la population réellement étudiée (appelée parfois *population-source*).

Cette différence peut être due à une différence entre la population cible et la base de sondage ou liste utilisée pour sélectionner l'échantillon. En théorie, la base de sondage est la liste exhaustive, sans omission ni répétition, des individus de la population. En pratique, les bases de sondage sont la plupart du temps imparfaites. L'erreur de couverture résulte de la différence créée sur les résultats de l'enquête par le fait de sélectionner l'échantillon à partir de cette liste imparfaite.

Plusieurs illustrations de cette erreur de couverture peuvent être données.

##### *Exemple 1*

Citons l'exemple des « listes rouges » et des « exclusifs portables » dans les enquêtes téléphoniques.

Les détenteurs exclusifs de téléphones mobiles ne représentaient en 1998 que 2 % de la population ; ils étaient environ 15 % en 2006, ce qui correspond à 17 % des ménages. Or, l'absence de ligne fixe s'avère lié avec un certain nombre de critères tels que l'âge, la profession et catégorie sociale (PCS) ou le niveau de revenu.

Pour limiter l'erreur de couverture des enquêtes par téléphone, certains instituts incluent des « individus exclusifs mobiles » ; une méthode possible est de générer aléatoirement des numéros de téléphonie mobile en respectant le poids des opérateurs de téléphonie mobile, d'identifier par un appel téléphonique parmi les numéros générés ceux qui correspondent à un individu « exclusif

mobile » et qui acceptent d'être enquêtés dans une étude ultérieure, et enfin de tirer un échantillon dans la base ainsi constituée (Dudoignon et Vanheuverzwyn, 2006)

Dans certaines enquêtes téléphoniques, des foyers sur listes rouges (environ 20% des foyers en France sont sur liste rouge) sont intégrés dans l'échantillon par la technique dite de « déclinaison ». Un tirage aléatoire de numéros de téléphone à partir de l'annuaire des abonnés au téléphone fixe est réalisé puis l'une ou l'autre des deux techniques suivantes est utilisée :

- exploitation des numéros du tirage initial puis déclinaison des numéros n'ayant pas abouti à une interview
- déclinaison immédiate des numéros du tirage initial et déclinaison des numéros n'ayant pas abouti à une interview.

Lorsque l'échantillon de numéros de téléphone est tiré sur l'annuaire France Telecom, la deuxième méthode permet en outre de joindre des ménages en liste rouge ou en dégroupage partiel

Dans une enquête téléphonique, la non-inclusion d'individus appartenant à l'une ou l'autre de ces deux sous-populations peut créer un biais dans les résultats si le comportement de ces individus diffère sur le sujet étudié.

On appelle *biais de sélection* le fait qu'en raison de non adéquation de la base de sondage et/ou des procédés de choix des unités échantillonnées, certains éléments ou groupes d'éléments sont exclus a priori. Chaque enquête comporte ses risques de biais de sélection et il faut y être très attentif; dans les enquêtes par quotas, par exemple, les consignes données aux enquêteurs ont pour but de diminuer ces risques.

Une autre cause de l'erreur de couverture est liée à des contraintes de réalisation de l'enquête sur le terrain : dans les enquêtes de satisfaction en vol des compagnies aériennes, les passagers des vols de nuit ne sont pas enquêtés ; dans les panels d'audience de la télévision, certains types de téléviseurs ne peuvent pas être équipés d'audimètres, etc.

Le biais résultant de l'erreur de couverture est la plupart du temps appréciée qualitativement.

### *Exemple 2*

Un autre exemple d'erreur de couverture est donné par les enquêtes Sans Domicile Fixe. L'absence de base de sondage fiable a été palliée par le recours aux structures proposant des services (restauration, hébergement, soin, etc.) pour constituer une base de sondage (Quaglia et Vivier (2006)). Ces enquêtes auprès de personnes sans domicile sont en fait plus précisément des enquêtes auprès des *utilisateurs* des services d'aide : les personnes qui ne fréquentent pas les structures échappent de fait à l'enquête.

Précisons que la méthode du partage des poids (Lavallée, 2002) a été utilisée pour limiter la surreprésentation des individus appartenant à plusieurs listes car fréquentant plusieurs centres d'accueil.

#### 4.2.2 L'erreur de non-réponse

L'erreur de non-réponse provient de l'absence totale ou partielle d'informations concernant des individus de l'échantillon. La tendance mondiale est à une augmentation du taux de non-réponse dans les enquêtes.

On distingue :

- la *non-réponse totale* qui peut être due à diverses causes : absence au moment du passage de l'enquêteur ou de l'appel téléphonique, refus de répondre, abandon en cours d'enquête, incapacité de participer à l'enquête, questionnaire annulé,
- la *non-réponse partielle* qui est due au refus de répondre devant certaines questions sensibles, à l'incompréhension du sens de la question, à des erreurs manifestes de codage ou de saisie conduisant à l'annulation de la réponse, etc.

Le problème dû aux *non-réponses partielles* est généralement moins aigu : les réponses aux autres questions du questionnaire donnent souvent des pistes d'explication exploitées par les méthodes *d'imputation* que nous décrirons plus loin.

Dans le cas où l'on fait l'hypothèse d'un modèle simple (voir Ardilly (2006) pour l'effet de la non-réponse sous d'autres modèles) dans lequel la population se compose de deux groupes : le groupe des répondants et le groupe des non-répondants, on montre que les réponses manquantes dans une enquête créent un *biais* qui dépend à la fois du taux de non-réponse *et* de l'écart entre les comportements des répondants et des non-répondants en ce qui concerne la ou les variables étudiées. Bien sûr, le comportement des non-répondants n'est pas connu ; la différence entre répondants et non-répondants peut parfois être appréciée qualitativement lorsque l'on dispose d'informations, de type sociodémographique par exemple, sur les non-répondants.

La non-réponse entraîne également une perte de précision puisque les résultats seront calculés sur un nombre de répondants plus faible que la taille d'échantillon initialement prévue. Cette perte de précision possible peut être anticipée en estimant *a priori* un taux de non-réponse probable et en augmentant en conséquence le nombre d'individus sélectionnés.

Le taux de non-réponse ne mesure donc pas à lui tout seul l'erreur de non-réponse : un taux de non-réponse faible peut entraîner un biais important si répondants et non-répondants ont des comportements très différents en ce qui concerne les thèmes de l'enquête. Inversement, un taux de non-réponse élevé n'est pas trop grave si ces comportements sont très voisins. Notons que ce biais ne diminue pas avec l'augmentation de la taille d'échantillon.

Le problème des non-réponses totales se pose différemment selon le type d'enquêtes et de méthodes de sondage utilisées :

- dans les méthodes non aléatoires (méthode des quotas par exemple lorsque l'enquête est réalisée en face à face), l'utilisateur de résultats n'a généralement pas connaissance du problème des non-réponses, des refus que l'enquêteur a subis, etc. Il peut éventuellement en avoir une appréciation sommaire, en examinant les consignes données aux enquêteurs et le déroulement du terrain ;
- dans les méthodes aléatoires, par contre, les individus composant l'échantillon sont nommément désignés. On connaîtra donc après enquête le nombre de non-réponses, éventuellement classées par catégorie ;

- dans les enquêtes en ligne, avec questionnaire proposé aux visiteurs de sites, l'utilisateur n'a généralement aucune connaissance des taux de réponse ;
- enfin, dans les panels, les non-répondants à une vague de panel sont bien identifiés ; on possède généralement beaucoup d'informations les concernant (via le questionnaire de recrutement et leurs réponses à d'autres vagues du panel). On peut alors utiliser les informations pour corriger l'effet des non-réponses ou pour estimer les réponses manquantes. Rappelons que les panels sont des échantillons permanents ou quasi-permanents d'individus ou de ménages qui sont interrogés ou observés régulièrement sur leurs comportements.

Deux types de stratégie peuvent être utilisés pour réduire l'erreur de non-réponse :

- la première stratégie consiste à réduire au maximum le taux de non-réponse en soignant très consciencieusement les techniques de collecte (questionnaires, enquêteurs, rappels...);
- la deuxième stratégie consiste à réduire *a posteriori* l'erreur de non-réponse en utilisant des techniques de repondération (pour la non-réponse totale) ou d'imputation (pour la non-réponse partielle).

#### a) Calcul des taux de réponse

Il n'existe pas de définition universellement acceptée des taux de réponse. Il faut donc être vigilant lorsque l'on analyse des taux de réponse ou dans leur comparaison.

Des efforts ont cependant été entrepris pour déterminer des règles de calcul : l'American Association for Public Opinion Research (AAPOR) a ainsi publié (AAPOR, 2008) un guide de calcul des taux de réponse dans les enquêtes par téléphone, dans les enquêtes en face-à-face à domicile et dans les enquêtes par voie postale.

De façon résumée,

le **taux de réponse** est défini comme le rapport du nombre d'interviews réalisés au nombre d'individus sélectionnées éligibles,

le **taux de refus** est défini comme le rapport du nombre de refus et d'abandons au nombre d'individus sélectionnées éligibles.

Pour ces deux derniers taux, plusieurs définitions sont possibles selon la façon dont on traite les cas d'éligibilité inconnue (par exemple, dans une enquête par téléphone, on ne sait pas si les numéros de téléphone qui ne répondent pas sont attribués ou non).

le **taux de coopération** est le rapport du nombre d'interviews réalisés au nombre d'individus éligibles joints.

le **taux de contact** est le rapport du nombre d'individus éligibles joints au nombre d'individus sélectionnées éligibles. C'est donc le rapport du total « interviews plus refus plus abandons » au nombre d'individus sélectionnées éligibles.

Le taux de réponse est égal au taux de coopération multiplié par le taux de contact.

*Exemple : Répartition des numéros de téléphone selon le résultat obtenu après le dernier appel sur la base des numéros utiles*

A titre d'illustration, on donnera ci-dessous une analyse du taux de réponse dans l'étude d'audience de la Presse Quotidienne et de la Presse Hebdomadaire Régionale réalisée depuis 2005 par TNS Sofres pour le compte d'Audipresse. Cette étude est réalisée par téléphone.

Jusqu'à fin 2007, un tirage initial de numéros était effectué à partir de l'annuaire téléphonique en respectant la matrice catégorie d'habitat x arrondissement, à raison de cinq numéros pour une interview à réaliser. Depuis janvier 2008, les numéros issus de l'annuaire téléphonique sont déclinés immédiatement (on ajoute une unité au dernier chiffre du numéro). Cette procédure permet d'inclure dans l'échantillon des numéros de ligne fixe en dégroupage partiel et de mieux représenter les individus dont le numéro figure sur liste rouge.

Les appels sont classés en fonction du résultat obtenu selon les modalités suivantes :

- |  |                                      |
|--|--------------------------------------|
| 1. interview acceptée                                | 7. faux numéro (numéro non attribué) |
| 2. abandon en cours d'interview                      | 8. rappel un autre jour              |
| 3. refus   | 9. prise de rendez-vous              |
| 4. hors champ (entreprise, fax, dialogue impossible) | 10. numéro occupé                    |
| 5. hors zone initiale                                | 11. sans réponse                     |
| 6. non-résident                                      |                                      |

Dans les quatre derniers cas, le numéro fait l'objet de dix tentatives d'appel. Si l'interview n'est pas obtenue après dix tentatives d'appel, le numéro est alors automatiquement décliné par le logiciel.

Les résultats suivants concernent l'étude réalisée entre juillet 2007 à juin 2008 (période du 25 juillet au 26 août exclue) et sont calculés sur la base des numéros utiles (numéros donnant lieu à une interview acceptée, un refus, un abandon, ou une « sans réponse », à l'exclusion des autres catégories). Le taux de réponse ainsi calculé de 22,9% ne tient pas compte du fait que certains numéros parmi les « sans réponse » ne sont pas attribués.

	Nombre de n°	%
INTERVIEW REALISEE	25 882	22,9
REFUS	46 835	41,5
ABANDON	3 053	2,7
SANS REPONSE	37 077	32,9
TOTAL	112 847	100

*Source : EPIQ 2007/2008*

La mesure et le suivi des taux de réponse dans les enquêtes répétitives permettent de surveiller un indicateur important de qualité mais aussi d'apprécier les difficultés de réalisation de

l'enquête sur le terrain, de pouvoir modifier en conséquence les consignes et les horaires d'interview et de suivre l'impact de ces modifications.

Lorsque les taux de réponse sont disponibles, on constate qu'ils diffèrent beaucoup :

- selon le thème de l'enquête : les enquêtes concernant la santé obtiennent généralement des taux de réponse plus élevés que les enquêtes sur l'audience des médias par exemple,
- la nature publique ou privée de l'enquête : l'INSEE obtient des taux de réponse à ses enquêtes qui sont pour la plupart supérieurs à 70% (Ardilly, 2006),
- le mode de collecte : les enquêtes en face-à-face sur base d'adresses ou par téléphone obtiennent généralement un meilleur taux de réponse que les enquêtes par Internet.

Il n'existe pas de réponse claire à la question : le taux de réponse est-il suffisant ? Le taux de réponse à une enquête doit être évalué par comparaison aux taux obtenus dans des enquêtes portant sur des thèmes semblables, avec des modes d'enquête comparables (cf. Brignier et Dupont, 2005).

#### *b) Les facteurs qui influent sur le taux de réponse*

Comme on l'a vu précédemment, le taux de non-réponse est une des deux composantes du biais de non-réponse. Il est évidemment plus facile de réduire ce taux que de réduire l'écart entre répondants et non-répondants sur les variables de l'étude. C'est la raison pour laquelle beaucoup d'efforts ont été consacrés à la compréhension des facteurs expliquant la non-réponse et au développement de techniques d'enquête permettant la réduction du taux de non-réponse (Biemer et Lyberg, 2003).

Pour réduire ce taux, il est indispensable :

- d'adapter le questionnaire, sa forme et sa longueur, au mode de collecte ; l'individu enquêté doit pouvoir facilement répondre au questionnaire ;
- de former les enquêteurs à la prise de contact avec les individus sélectionnés afin d'éviter au maximum les refus ou les abandons,
- de faire en sorte d'obtenir une réponse des individus sélectionnés par des visites (enquêtes à domicile sur adresses, des rappels (enquêtes par téléphone) ou des relances (voie postale ou enquêtes sur base d'adresses e-mail)

D'autre part, une récompense ou rémunération accordée aux individus qui acceptent de répondre accroît généralement le taux de réponse.

#### *c) Le traitement de la non-réponse*

*A posteriori*, il est possible de réduire l'effet de la non-réponse par l'utilisation de méthodes statistiques adaptées.

*Pour la non-réponse totale*, une stratégie envisageable est celle de la *repondération* de l'échantillon des répondants : on réajuste sur des distributions connues. Formellement, les calculs sont les mêmes que ceux de la post-stratification ou du redressement sur critères multiples, mais ils n'ont pas le même statut théorique. Dans ces dernières méthodes, on considère que les écarts entre les distributions observées sur l'échantillon et sur la population résultent des seuls aléas

d'échantillonnage et les calculs relèvent de la seule théorie mathématique de l'estimation. Dans la repondération pour non-réponse, on postule, de plus, l'hypothèse plus ou moins vérifiable que, dans chaque catégorie définie par les variables de redressement, répondants et non-répondants sont en moyenne semblables. Les méthodes de repondération permettent de réduire le biais dû à la non-réponse mais pas nécessairement la variance.

*Pour la non-réponse partielle*, on utilise des méthodes par *imputation* (Ardilly (2006)) : on remplace chaque donnée manquante par une donnée « prédite » en fonction des informations obtenues pour le même individu et pour des individus proches. On peut distinguer deux grandes classes de méthodes d'imputation :

- l'imputation *déterministe* est telle que la valeur imputée ne change pas si l'on relance la procédure d'imputation. Parmi les méthodes, citons l'imputation utilisant un modèle explicite de type linéaire : les coefficients du modèle sont estimés sur l'échantillon des répondants et permettent d'estimer les valeurs imputées pour les non-répondants.
- l'imputation *aléatoire* est telle que la valeur imputée change si l'on relance la procédure d'imputation. Parmi les méthodes possibles, citons les méthodes de hot-deck dans lesquelles la valeur inconnue de la variable d'intérêt pour un non-répondant est remplacée par la valeur observée d'un répondant tiré au hasard, soit parmi l'ensemble des répondants (hot-deck d'ensemble), soit dans une classe à laquelle appartient le receveur (hot-deck par classe), etc.

#### **4.2.3. L'erreur de mesure**

Le processus de mesure est certainement la source d'erreur la plus complexe dans une enquête par sondage. Il comprend six composantes de base : le mode de collecte des données, l'enquêteur, l'interviewé, le questionnaire, le système d'information et l'environnement. Le système d'information fait référence à l'ensemble des informations dont l'interviewé dispose pour formuler une réponse ; il peut par exemple consulter des documents, ou d'autres personnes dans son foyer ou sa propre mémoire. L'environnement concerne le lieu dans lequel a lieu l'interview : ce peut être le domicile des interviewés, une classe, un hôpital, un magasin, etc. (Biemer et Lyberg (2006)).

Toutes ces composantes interagissent durant le processus de mesure de façon complexe ; par exemple, une formulation de question peut avoir différents effets sur la réponse selon le mode d'administration du questionnaire, les caractéristiques de l'enquêteur, la nature de la population enquêtée, etc.

Comme nous l'avons écrit précédemment, *l'erreur de mesure* (ou d'observation) provient d'écarts entre les réponses enregistrées et les vraies valeurs.

Elle peut être due essentiellement :

- à l'instrument de mesure (mode de collecte, questionnaire,...),
- à l'interviewé (il fournit une information incorrecte, volontairement ou involontairement),
- aux enquêteurs (d'où formation et contrôle nécessaires).

### a) *L'effet enquêteur*

Nous renvoyons le lecteur à l'ouvrage de Biemer et Lyberg (2006) pour une description détaillée de ce que l'on appelle souvent l'effet enquêteur.

Le « Guide pratique de la qualité en études de marché » publié par Syntec – Etudes de marché insiste sur l'importance de la réalisation de l'étude sur le terrain ; cette réalisation est encadrée par trois temps forts : l'entraînement des enquêteurs ou *briefing*, point de départ indispensable, l'encadrement des enquêteurs sur le terrain par accompagnement ou écoutes téléphoniques, enfin leur contrôle.

### b) *L'influence du mode de collecte des informations*

L'influence du mode de collecte des informations sur la qualité des résultats est aussi décrite en détail dans de nombreux ouvrages (par exemple, Biemer et Lyberg (2006), dans le contexte général des enquêtes par sondage, Evrard et *al.* (2003), Jolibert et Jourdan (2006) dans le contexte des études et recherche en marketing) et ne sera pas détaillée dans ce document.

Signalons toutefois que le développement des enquêtes assistées par ordinateur (système CATI pour « Computer Assisted Telephone Interviewing », système CAPI pour « Computer Assisted Personal Interviewing », système CAWI pour « Computer Assisted Web Interview » etc.) a généralement contribué à l'amélioration de la qualité des enquêtes. En effet, ces systèmes

- offrent la possibilité dans les enquêtes par téléphone de rappeler les numéros absents, de gérer des rendez-vous téléphoniques,
- favorisent une plus grande homogénéité du travail des enquêteurs,
- permettent une permutation aléatoire des items, diminuant l'effet d'ordre qui sera mentionné ci-dessous,
- permettent l'inclusion de tops horaires pour mesurer les durées d'administration des différentes parties du questionnaire et contrôler ainsi le travail des enquêteurs.

### c) *Le questionnaire*

S'il est possible d'identifier les spécificités des différents modes de collecte des informations, il est plus difficile de donner des règles précises pour la construction d'un bon questionnaire : le savoir-faire et l'expérience y jouent un rôle déterminant.

Notre objectif dans ce paragraphe sera donc essentiellement de sensibiliser le lecteur à l'influence de la formulation des questions et de la construction du questionnaire sur les résultats obtenus. Ce paragraphe reprend largement des règles exposées dans Dussaix et Grosbras (1996). Le lecteur trouvera de nombreux ouvrages en bibliographie sur la formulation des questions et leur impact sur les réponses.

Comme nous l'avons déjà dit, la mise au point du questionnaire comprend généralement les étapes suivantes :

- étude qualitative ou exploratoire (par entretiens en profondeur, entretiens de groupe) ; leur objectif est de contribuer à la définition et à la formulation du problème, d'identifier les dimensions pertinentes (ex. : les composantes de la qualité dans une étude sur la qualité de

service d'une chaîne de restaurants), d'élaborer des hypothèses, d'identifier les mots et expressions utilisés par la population étudiée,

- première rédaction du questionnaire,
- pilotage (ou prétest) du questionnaire auprès d'un petit échantillon reflétant la diversité de la population visée (entre 10 et 30 individus) : cette phase est indispensable et permet de s'assurer que les questions et les mots utilisés sont bien compris, que le questionnaire est fluide et n'est pas trop long,
- version définitive du questionnaire en tenant compte des observations formulées par les enquêteurs et les enquêtés dans la phase précédente

### c.1) Le format des questions

On distingue essentiellement les questions *ouvertes* et les questions *fermées*.

Une question *ouverte* est une question à laquelle l'interviewé répond comme il le désire ; ce qu'il dit est en général intégralement enregistré par l'enquêteur

*Exemple:*

A quoi principalement, occupez-vous votre temps libre en dehors des vacances?

Une question *fermée* est une question à laquelle les réponses possibles sont déterminées à l'avance par le rédacteur du questionnaire, proposées à l'enquêté souvent sur un carton-réponse en lui précisant le nombre de réponses qu'il peut choisir.

*Exemple :*

Quel(s) moyen(s) de transport avez-vous utilisé(s) pour vous rendre sur votre principal lieu de vacances ?

- train
- avion
- voiture de location
- voiture personnelle
- autre

Les questions ouvertes permettent de recueillir une information parfois plus nuancée que les questions fermées, donnent à l'interviewé le sentiment que son avis compte, permettent d'aérer un questionnaire complexe. On essaie cependant d'éviter les questions ouvertes dans les études quantitatives ; on y obtient souvent des réponses superficielles ou stéréotypées, l'interviewé citant les premières raisons qui lui viennent à l'esprit; d'autre part, l'analyse des questions ouvertes s'avère lourde et coûteuse malgré l'existence de logiciels « ad hoc » : il faut en effet regrouper les réponses ayant une signification semblable en un nombre limité de thèmes, ce qui implique de surcroît une certaine part de subjectivité. Mieux vaut donc « fermer » les questions en utilisant les résultats d'une étude qualitative préalable.

De très nombreuses formulations de questions ont été proposées pour la mesure des opinions et des attitudes (exemples : échelle de Thurstone, échelle de Likert, échelle sémantique différentielle). Nous reportons le lecteur à des ouvrages spécialisé pour un exposé de ces différents types d'échelle, de leurs avantages et inconvénients (Evrard Y. et al. (2003), Jolibert et Jourdan (2006)).

### c.2) La formulation des questions

La formulation des questions doit respecter des règles simples et essentielles.

*Première règle* : les questions doivent être *comprises* par l'interviewé.

Il faut donc éviter (liste non exhaustive)

- *le vocabulaire technique* lorsque le questionnaire s'adresse au grand public. Cette recommandation est particulièrement importante dans les enquêtes portant sur les nouvelles technologies : modes de réception de la télévision par exemple. Il faut essayer au maximum d'adapter la formulation des questions au vocabulaire utilisé par les interviewés,
- *les mots dont le sens est vague* : « lire un magazine » n'a pas la même signification pour tous. On préfère utiliser la formulation extensive :

« Personnellement, avez-vous lu, parcouru ou consulté... ? »

Il en est de même des termes « souvent, quelquefois, rarement ». Il vaut mieux poser :

Pour chacun des journaux gratuits que vous avez mentionnés, indiquez la fréquence à laquelle vous avez l'habitude de les lire, les parcourir ou les consulter :

- |                     |                          |
|---------------------|--------------------------|
| toutes les semaines | <input type="checkbox"/> |
| 2 à 3 fois par mois | <input type="checkbox"/> |
| 1 fois par mois     | <input type="checkbox"/> |
| 5 à 6 fois par an   | <input type="checkbox"/> |
| moins souvent       | <input type="checkbox"/> |

- *les doubles négations*

Une question telle que :

« Ne pensez-vous pas qu'on ne doit pas autoriser l'ouverture des grandes surfaces le dimanche ?

demande à l'interviewé un certain entraînement...

- *les questions comportant plusieurs idées à la fois*. En voici un exemple :

La ville de Victus entend continuer à favoriser le développement équilibré de son parc de logements aidés en participant au financement

- des logements PLA (ex HLM) pour les revenus les plus modestes
- des logements intermédiaires pour les classes moyennes
- Qu'en pensez-vous?
  - Je suis d'accord
  - Je ne suis pas d'accord
  - Sans opinion

Cette question est un modèle du genre à éviter ; elle comporte deux questions : que répond l'enquêteur qui est pour le financement des logements PLA mais contre celui des logements

intermédiaires ; que veut dire « logements intermédiaires »? Qui serait contre un développement équilibré?

*Deuxième règle* : les interviewés doivent être *capables* de répondre à la question. Il faut donc en particulier s'assurer

- que, lorsqu'une liste de réponses possibles est proposée, toutes les situations possibles sont bien prévues;
- que la réponse n'est pas trop difficile pour la mémoire ; dans certaines études (exemple : détention des produits financiers par les ménages) le questionnaire auto-administré peut être plus adapté que le questionnaire par enquêteur en face à face : il permet la consultation des documents par l'interviewé.

*Troisième règle* : la formulation des questions doit permettre d'obtenir des réponses *sincères*.

Il faut éviter les questions biaisées dont la formulation induit les réponses ; qui ne serait pas d'accord avec la proposition suivante ?

La municipalité a l'ambition de faire de Victus la capitale économique de l'Europe de demain.

Qu'en pensez-vous?

- Je suis d'accord
- Je ne suis pas d'accord
- Sans opinion

Il faut aussi penser :

- au *biais de désirabilité sociale*, aux réactions de prestige de l'interviewé d'où des sous-estimations des consommations d'alcool mais des surestimations des consommations de shampoing ou de dentifrice
- à la peur du changement ou tendance au conformisme
- à *l'attraction de la réponse positive* ou tendance des interviewés à répondre oui, vrai, d'accord, etc.

Enfin, dans l'interprétation des résultats, il ne faut pas oublier que l'*ordre de présentation des réponses* a un effet sur les résultats comme le souligne l'exemple suivant (Grémy, dans Lebart, 1992) dans lequel un questionnaire a été présenté sous deux formes, chacune proposant dans un ordre différent la liste des réponses possibles à la question : « Selon vous, quels sont aujourd'hui les deux problèmes les plus graves ? : le premier plus grave, le second plus grave ? ».

Le tableau ci-dessous donne les pourcentages de réponse pour le problème jugé le plus grave. On remarque un effet assez sensible de l'ordre de présentation (même si les trois problèmes les plus graves sont identiques).

Problème le plus grave aujourd'hui		
Réponse	Ordre : direct n <sub>1</sub> = 1186	Ordre : indirect n <sub>2</sub> = 1186
le chômage	34,2%	19,2%
le terrorisme	8,5%	7,0%
la faim dans le monde	21,3%	17,7%
la guerre	15,3%	18,3%
la superpopulation du monde	1,5%	1,7%
le racisme	5,0%	8,5%
le non-respect des droits de l'homme	8,0%	10,4%
l'insuffisante formation professionnelle des jeunes	1,3%	6,2%
la délinquance	1,2%	8,2%

Dans les études d'audience, par exemple l'étude EPIQ d'audience de la Presse Quotidienne et de la Presse Hebdomadaire Régionale, les titres sont présentés aux interviewés après permutation aléatoire des titres à l'intérieur de chaque groupe de titres (Presse Quotidienne Régionale, Presse Quotidienne Nationale, Presse Quotidienne Gratuite d'Information, etc.) afin de limiter l'effet d'ordre.

Il ne faut pas oublier non plus dans l'interprétation et dans la comparaison de résultats l'importance du *choix des mots*.

### c.3) L'ordre des questions dans le questionnaire

L'ordre des questions dans le questionnaire a une influence sur les réponses, que ce soit :

- par *effet de halo* : lorsque les questions sont des échelles orientées dans le même sens (ex. : échelles de satisfaction portant sur la qualité des plats, le rapport qualité-prix, l'accueil), les interviewés ont tendance à répondre de manière semblable :
- par *effet de contamination*, la suite des questions posées ayant une influence sur les réponses aux questions suivantes.

Il vaut mieux commencer le questionnaire par des questions d'ordre général qui introduisent le sujet, suscitent la motivation de l'interviewé, et terminer par les questions d'identification (âge, profession)

### c.4) La longueur du questionnaire

La durée acceptable d'un questionnaire dépend de plusieurs facteurs

- le mode d'administration : un questionnaire par téléphone doit être plus court qu'un questionnaire administré par enquêteur en face-à-face à domicile,
- l'endroit où il est administré : un questionnaire posé « dans la rue » ne devrait pas dépasser 10 à 15 minutes ; on donne généralement une limite supérieure de 45 minutes s'il est administré à domicile,
- et l'intérêt du sujet pour les interviewés.

Citons l'enquête nationale Analyse des Comportements Sexuels des Français (ACSF) de l'INSERM dans le contexte de l'épidémie du SIDA (Spira et Bajos, 1993). Un test comparant face à

face et téléphone a montré que l'enquête pouvait être réalisée par téléphone avec un questionnaire d'une durée moyenne de 45 minutes, alors qu'on admet généralement qu'une enquête par téléphone ne devrait pas dépasser 15 à 30 minutes.

### 4.3. L'erreur totale

L'erreur totale dans une enquête est la somme de l'erreur d'échantillonnage et des autres types d'erreur qui ont été explicités dans les paragraphes précédents.

Dans la conception d'une enquête, il faut donc tenter de choisir la taille d'échantillon, le mode d'administration du questionnaire, le réseau d'enquêteurs, les procédures de contact et de relance des interviewés, etc., qui minimisent l'erreur totale en tenant compte des contraintes de coût et de délais.

La mesure de l'erreur totale la plus souvent utilisée dans la littérature concernant les enquêtes par sondage est l'erreur quadratique moyenne ou EQM (*Mean squared error* ou *MSE*).

La conception d'une enquête va donc consister à minimiser l'erreur quadratique moyenne des estimateurs des variables d'intérêt tout en respectant les contraintes de budget et de délais.

Dans ce contexte, une classification utile des erreurs non d'échantillonnage consiste à distinguer :

- les erreurs variables
- les erreurs systématiques

Les *erreurs variables* sont des erreurs qui se compensent entre les répondants ; en prenant l'exemple d'une question sur la durée d'écoute de la télévision par les interviewés la veille de l'interview, on parle d'erreur variable si les interviewés ont autant de chances de déclarer une durée d'écoute supérieure à la réalité qu'une durée d'écoute inférieure.

Par contre, on parlera d'*erreur systématique* ou biais si les erreurs positives et négatives ne s'annulent pas ; ce sera le cas si les interviewés sont conduits par la formulation de la question à déclarer une durée d'écoute supérieure à ce qu'elle est réellement.

Les erreurs variables sont moins dommageables pour la qualité des résultats, en ce sens que leur effet sur les estimations des paramètres habituellement mesurés dans les enquêtes (moyennes, totaux, proportions) diminue si l'on augmente la taille d'échantillon. Ce n'est pas le cas des erreurs systématiques.

Nous allons définir l'erreur quadratique moyenne (EQM) sans introduire de formalisme statistique. Pour un paramètre caractéristique de la population, l'EQM va être définie comme la moyenne des écarts au carré entre les valeurs  $\hat{T}$  obtenues dans de nombreuses répétitions hypothétiques du processus d'enquête et la vraie valeur inconnue  $T$  du paramètre.

Par analogie avec l'EQM d'un estimateur en statistique, on peut écrire :

$$EQM = E(\hat{T} - T)^2$$

en notant  $E$  la moyenne sur toutes les répétitions hypothétiques de l'enquête.

On peut également écrire :

$$EQM = [E(\hat{T}) - T]^2 + E[\hat{T} - E(\hat{T})]^2$$

Soit : 
$$EQM = (\text{Biais})^2 + \text{Variance}$$

Le biais  $[E(\hat{T}) - T]$  est non nul si la moyenne des valeurs obtenues sur toutes les répétitions possibles de l'enquête diffère de la vraie valeur inconnue. En prenant l'analogie du tir sur cible, s'il y a biais, les impacts des tirs ne sont pas répartis autour du cœur de la cible.

La variance  $E[\hat{T} - E(\hat{T})]^2$  traduit la dispersion des valeurs obtenues dans toutes les répétitions possibles de l'enquête autour de leur moyenne  $E(\hat{T})$  qui, s'il y a biais, diffère de la vraie valeur inconnue T.

Comme nous l'avons déjà mentionné, l'erreur d'échantillonnage contribue essentiellement à la variance et diminue lorsque la taille d'échantillon augmente ; l'erreur de couverture et l'erreur de non-réponse contribuent essentiellement au biais. L'erreur de non-réponse peut contribuer également à augmenter la variance si les répondants sont affectés de coefficients de redressement élevés par l'utilisation de méthodes de pondération pour corriger le biais de non-réponse ; il en résulte un arbitrage à faire entre réduction du biais et réduction de la variance.

Cette décomposition de l'erreur quadratique moyenne en une somme du biais au carré et de la variance sert de *modèle* pour arbitrer entre différents schémas d'enquête lors de la conception d'une enquête mais ne permet généralement pas de calculer explicitement les différentes composantes du biais et de la variance. L'évaluation de la qualité sera donc réalisée à travers un examen approfondi de la méthodologie d'enquête et de la réalisation de l'enquête sur le terrain, comme nous le précisons au paragraphe 5. cependant, cette décomposition met en évidence le fait qu'une grande taille d'échantillon ne suffit pas pour obtenir des résultats fiables si le processus d'enquête (mauvais plan de sondage, mauvais contrôle des enquêteurs, mauvais questionnaire,...) génère des biais importants.

## 5. L'évaluation de la qualité

L'évaluation de la qualité dans les enquêtes a pour objectifs :

- d'identifier les sources d'erreur et de corriger l'impact de ces erreurs
- de faire progresser cette qualité au cours de la réalisation de l'enquête ou dans une enquête ultérieure
- de fournir aux utilisateurs de données une évaluation de la qualité des résultats d'enquête.

Ce processus d'évaluation peut être conduit lors de chacune des étapes de l'enquête, par exemple :

- études pilote pour évaluer des schémas d'enquête alternatifs lors de la conception du questionnaire,

- pré-test du questionnaire pour s'assurer que les questions et les mots utilisés sont bien compris, que le questionnaire est fluide et n'est pas trop long,
- identification des sources d'erreur durant le déroulement de l'enquête afin de les corriger s'il en est encore temps (exemples : participation aux briefings des enquêteurs, contrôle du respect des consignes et du bon déroulement du questionnaire à l'aide d'écoutes des interviews dans une enquête par téléphone, etc.). Ces contrôles permettent aussi de vérifier que le cahier des charges de l'enquête est bien suivi (Giudicelli et al. (2006) décrit de façon détaillée les points à évaluer dans une enquête par téléphone).

Les audits du Centre d'étude des supports de publicité (CESP) constituent un exemple de ce type d'évaluation ; ils concernent les études « de référence » sur la mesure d'audience des médias (Brignier et al. ,2002) et s'exercent à un double niveau, en distinguant :

- le contrôle de la réalisation de l'étude qui est une synthèse des enseignements tirés des contrôles relatifs au déroulement de l'enquête,
- l'audit de la méthodologie de l'enquête dans lequel la cohérence entre les choix méthodologiques effectués et les objectifs de l'enquête est examinée (plan de sondage, mode d'interview, questionnaire, heures et durée de l'interview, structure de l'échantillon observé et méthode de redressement, analyse des résultats et de leur cohérence).

## Ouvrages et articles cités dans l'article :

AAPOR (2008), *Standard Definitions : Final Dispositions for Case Codes and Outcome Rates for Surveys*, The American Association for Public Opinion Research.

Ardilly, P. (2006). *Les techniques de sondage*, Technip.

Biemer, P.P., Lyberg L.E. (2003). *Introduction to Survey Quality*, Wiley Series in Survey Methodology.

Brackstone, G.J. (1999). *La gestion de la qualité des données dans un bureau de statistique*, Techniques d'Enquête, Vol. 25, n°2.

Brignier, J.M. et Dupont, F. (2005). Taux de réponse et qualité des enquêtes téléphoniques, *Décisions Marketing* n°38, p. 47-57.

Deming, W.E. (1944), On Errors in Surveys, *American Sociological Review*, 9, pp. 359-369.

Dudoignon, L. et Vanheuverzwyn, A. (2006). *Le téléphone mobile dans les enquêtes de référence de la mesure d'audience des médias en France*, dans *Méthodes d'enquêtes et sondages*, Lavallée, P. et Rivest, L.P., éditeurs, Paris, Dunod.

Dussaix, A.M., Grosbras, J.M. (1996). *Les sondages : principes et méthodes*, Que sais-je n°701, Presses Universitaires de France, 2<sup>ème</sup> édition.

EUROSTAT, *Handbook on Data Quality Assessment Methods and Tools*, Ehling, M. et Körner, T., (eds).

EUROSTAT (2005). *Standard Quality Indicators*, Working Group "Quality in statistics", Luxembourg, 23-24 May 2005.

- Evrard, Y., Pras, B. et Roux, E. (2006), *Market : études et recherche en marketing*, 3<sup>ème</sup> édition, Paris, Dunod.
- Giudicelli, E., Léon, C., Arwidson, P. et Guilbert, P. (2006). *La qualité des données dans les enquêtes par téléphone : le recours à une société de surveillance du terrain*, dans *Méthodes d'enquêtes et sondages*, Lavallée, P. et Rivest, L.P., éditeurs, Paris, Dunod.
- Jolibert, A., Jourdan, P. (2006). *Marketing Research*, Paris, Dunod..
- Lavallée, P. (2002), *Le sondage indirect ou la méthode généralisée de partage des poids*, Editions de l'Université de Bruxelles et Paris, Editions Ellipses.
- Lebart L. (éditeur), (1992). *La qualité de l'information dans les enquêtes*, Dunod.
- Quaglia, M. et Vivier, G. (2006), *Enquêter des populations difficiles à joindre : principes méthodologiques et adaptations pratiques*, dans *Méthodes d'enquêtes et sondages*, Lavallée, P. et Rivest, L.P., éditeurs, Paris, Dunod.
- Spira, A., et Bajos, N. (1993). *Les comportements sexuels en France*, Paris, La documentation française.
- SYNTEC – Etudes de marché, *Le Guide Pratique de la Qualité en Etudes de Marché*.
- Tillé, Y. (2001). *Théorie des sondages*, Dunod.

### **Autre ouvrages et articles :**

- Biemer, P.P., Groves, R.M., Lyberg, L.E., Mahiowetz, N.A. et Sudman, S. (1991). *Measurement errors in surveys*, Wiley, 1991.
- Brossier G., Dussaix A.M., éditeurs, (1999). *Enquêtes et sondages: Méthodes, modèles, applications ; nouvelles approches*, Dunod,
- Collins, M. and Sykes, W. (1999). Extending the Definition of Survey Quality, *Journal of Official Statistics*, vol. 15, N° 1, pp. 57-66.
- Dupont F. (1999), Etudes sur la mobilité et implication sur les plans de sondage, dans *Enquêtes et sondage : Méthodes, modèles, applications, nouvelles méthodes*, G.Brossier et A.M.Dussaix éditeurs, Dunod, 1999.
- Groves, R.M. (1989). *Survey Errors and Survey Costs*, Wiley.
- Groves, R.M. and Couper, M.P., (1998), *Nonresponse in Household Interview Surveys*, Wiley.
- Kish, L. (1987), *Statistical Design for Research*, Wiley,
- Lejeune M. (éditeur), (2001). *Traitements des fichiers d'enquêtes : redressements, injections de réponses, fusions*, Presses Universitaires de Grenoble.
- Lessler J.T., Kalsbeek W.D. (1992). *Nonsampling Errors in Surveys*, Wiley.
- Lyberg, L., Biemer, P., Collins, M., De Leeuw, E., Dippo, C., Schwarz, N., and Trewin, D. (eds.) (1997). *Survey Measurement and Process Quality*. New York: Wiley.
- Särndal, C.E., Swensson, B. et Wretman, J.H., (1992), *Model Assisted Survey Sampling*, Springer Verlag.
- Särndal, C.E. et Lundström, S. (2005). *Estimation in Surveys with Nonresponse*, Wiley.
- Sudman, S. and Bradburn, N. (1989). *Asking Questions*, Jossey-Bass Publishers.

Sudman, S., Bradburn N., and Schwarz, N. (1996). *Thinking about Answers : The Application of Cognitive Processes to Survey Methodology*. San Francisco : Jossey-Bass.

Tanur, J.M. (ed.) (1992). *Questions about Questions : Inquiries into the Cognitive Bases of Surveys*, Sage, pp154-169.

Tourangeau, R., Rips, L., and Rasinski, K. (2000). *The Psychology of Survey Response*. New York : Cambridge University Press.

**Autre sources :**

Le lecteur pourra consulter également :

- les actes des « Colloques francophones sur les sondages », édités par Dunod, respectivement en 1999, 2001, 2004, 2006, et 2008. Consulter également le site de la Société française de statistique (SFdS) pour une information sur ces colloques ([www.sfds.asso.fr](http://www.sfds.asso.fr)) ;
- les exposés du séminaire « La qualité dans les enquêtes » organisé en juin 2007 par le groupe Enquêtes de la SFdS et disponibles sur le site ([www.sfds.asso.fr](http://www.sfds.asso.fr)) ;

Des codes de déontologie et des publications peuvent être trouvés sur les sites de :

- Syntec Etudes Marketing et Opinion ([www.syntec-etudes.com](http://www.syntec-etudes.com)) : Code international ICC/ESOMAR des Etudes de Marché et d'Opinion),
- ESOMAR ([www.esomar.org](http://www.esomar.org)) : de nombreux codes professionnels sont accessibles, par exemple « Interviewing children and young people ».

Taille d'Echantillon	Proportion f						
	5 % ou 95 %	10 % ou 90 %	15 % ou 85 %	20 % ou 80 %	30 % ou 70 %	40 % ou 60 %	50%
100				7,8	9,0	9,6	9,8
150			5,7	6,4	7,3	7,8	8,0
200		4,2	4,9	5,5	6,4	6,8	6,9
300	2,5	3,4	4,0	4,5	5,2	5,5	5,7
400	2,1	2,9	3,5	3,9	4,5	4,8	4,9
500	1,9	2,6	3,1	3,5	4,0	4,3	4,4
700	1,6	2,2	2,6	3,0	3,4	3,6	3,7
1000	1,4	1,9	2,2	2,5	2,8	3,0	3,1
2000	1,0	1,3	1,6	1,8	2,0	2,1	2,2
5000	0,6	0,8	1,0	1,1	1,3	1,4	1,4
10000	0,4	0,6	0,7	0,8	0,9	1,0	1,0