Techniques d'enquête, décembre 1975
Vol. 1, nº 2
Statistique Canada, nº 12-001-X au catalogue

Mesure des erreurs de réponse dans les recensements et les enquêtes par sondage

G.J. Brackstone, J.F. Gosselin et B.E. Garton[1]

Résumé

Madow (l968) a proposé un schéma d'échantillonnage à deux degrés suivant lequel le biais de réponse peut être éliminé des enquêtes par sondage en obtenant des valeurs « réelles » pour un sous-échantillon de l'échantillon original. Comme c'est souvent le cas aux recensements ou aux enquêtes en cours, les données des sous-échantillons ne servent pas à corriger les estimations de l'enquête principale, mais à évaluer leur fiabilité. Ce document vise d'abord à présenter des méthodes permettant d'obtenir des estimations de fiabilité lorsque les valeurs « réelles » peuvent être établies pour un sous-échantillon d'unités.

Mots-clés : Sous-échantillon; estimations de fiabilité; échantillonnage à deux degrés.

1 Introduction

Madow (1965) a proposé un plan d'échantillonnage prévoyant l'élimination du biais de réponse des estimations d'enquête par sondage par l'obtention d'« observations vraies » tirées d'un sous-échantillon de l'échantillon initial, et ce, par l'application de l'estimation de biais de ce sous-échantillon pour corriger l'estimation initiale.

On note toutefois des cas où les données du sous-échantillon sont obtenues à des fins d'évaluation après la publication des données d'enquête. L'objectif consiste alors à mesurer la fiabilité générale des données d'enquête déjà publiées afin :

a) de renseigner l'utilisateur des données sur leur fiabilité, d'où la possibilité pour lui de les rajuster s'il le désire;
b) de renseigner l'auteur de l'enquête sur les sources d'erreur de cette dernière de manière à pouvoir améliorer les enquêtes futures.

Notre propos sera donc ici de présenter des estimateurs de la fiabilité des données de recensement ou d'enquête par sondage quand des valeurs « réelles » peuvent être tirées d'un échantillon ou d'un sous-échantillon de cas.

1. G.J. Brackstone, J.F. Gosselin et B.E. Garton, Division des méthodes d'enquêtes – recensement et enquêtes-ménages.

Nous abordons le problème sous un double angle. D'abord, nous prenons le cadre habituel de la variance et du biais de réponse. Ce sera là la principale partie de notre exposé. Ensuite, nous écartons cet angle d'attaque et tentons de mesurer seulement l'erreur nette dans l'enquête particulière que nous observons.

Nos résultats s'appliquent à tout plan d'échantillonnage de l'enquête initiale et à tout plan de sous-échantillonnage permettant d'obtenir des valeurs « réelles », à la seule condition que des estimateurs de la variance d'échantillonnage soient disponibles pour les plans d'échantillonnage en question.

2 Approche I

2.1 Modèle de l'erreur de réponse

Le modèle de l'erreur de réponse repose sur la notion de répétition indépendante. Parlons d'abord du cas du recensement (dénombrement à 100 %). Nous posons l'hypothèse que de nombreux « essais » indépendants du recensement pourraient se faire dans les mêmes conditions générales. L'« estimation » de recensement pour une catégorie d'intérêt quelconque suivrait alors une certaine distribution de fréquences. Un chiffre donné du recensement pourrait dans ce cas être considéré comme une observation aléatoire venant d'une distribution de toutes les estimations possibles de recensement.

Soit $\bar{X}(t)$ l'estimation de recensement obtenue à l'essai t et soit $\bar{\mu}$ la moyenne vraie correspondante. Le paramètre statistique par lequel on mesure habituellement la fiabilité d'une estimation en pareil cas est l'erreur quadratique moyenne (EQM) de $\bar{X}(t)$:

$$\begin{aligned} \text{EQM}\,(\bar{X}(t)) &= \underset{t}{E}\,(\bar{X}(t) - \bar{\mu})^2 \\ &= \underset{t}{V}\,[\bar{X}(t)] + B^2, \end{aligned}$$

où B désigne le biais de $\bar{X}(t)$,

c'est-à-dire $B = \underset{t}{E}\,[\bar{X}(t)] - \bar{\mu}$

$= \bar{X} - \bar{\mu}$ où $\bar{X} = \underset{t}{E}\,[\bar{X}(t)]$.

La variance de réponse de $\bar{X}(t)$ se définit comme

$$\underset{t}{V}\,(\bar{X}(t)) = \underset{t}{E}\,(\bar{X}(t) - \bar{X})^2 .$$

En général donc, il y a biais lorsque les erreurs se présentent ordinairement dans une direction plutôt que dans une autre. Si une erreur s'était glissée, par exemple, dans les instructions accompagnant le questionnaire de recensement, les erreurs tendraient à être systématiques en une direction seulement. Ainsi, le biais mesure l'effet net moyen de tous les facteurs susceptibles d'entrer en jeu.

Par ailleurs, la variance de réponse mesure la composante aléatoire de l'erreur. Si une question est ambiguë par exemple, quelqu'un qui répond en autodénombrement pourrait donner des réponses qui varient selon les essais indépendants. Ces types d'erreur dépendent de facteurs inconnus indépendants de notre volonté et qui varient d'un essai à l'autre (qu'il s'agisse de la mentalité du recensé ou de la fatigue de l'intervieweur).

Ce qui précède vaut pour toute caractéristique venant d'un recensement, mais on peut aisément l'étendre aux caractéristiques d'une enquête par sondage, auquel cas si $\bar{x}_s(t)$ désigne l'estimation venant de l'échantillon s à l'essai t, l'EQM $[\bar{x}_s(t)]$ peut revêtir la forme suivante :

$$EQM[\bar{x}_s(t)] = E\,[\bar{x}_s(t) - \bar{\mu}]^2 ,$$

où l'espérance se prend sur l'ensemble des essais et des échantillons possibles :

Soit $\bar{X} = E\,(\bar{x}_s(t)) = \underset{s}{E}\,\underset{t}{E}\,(\bar{x}_s(t)|s)$

et $B = \bar{X} - \bar{\mu}$.

$$EQM\,[\bar{x}_s(t)] = E\,(\bar{x}_s(t) - \bar{X})^2 + B^2$$

$$= \underset{s}{E}\,\underset{t}{V}\,(\bar{x}_s(t)|s) + \underset{s}{V}\,\underset{t}{E}\,(\bar{x}_s(t)|s) + B^2 ,$$

où le premier terme mesure la variance de réponse, le second, la variance d'échantillonnage, et le troisième, le biais.

2.2 Estimation

Dans le modèle de l'erreur de réponse que nous avons décrit, trois paramètres statistiques se trouvent définis, à savoir l'erreur quadratique moyenne, la variance de réponse et le biais. Idéalement, nous rechercherions des estimations de ces trois paramètres de manière à évaluer le niveau des erreurs tant aléatoires que systématiques et à dégager une mesure globale de la fiabilité.

Par l'hypothèse selon laquelle la valeur réelle d'une caractéristique de l'échantillon d'enquête est connue pour un sous-échantillon aléatoire de celui-ci, nous établissons des estimations sans biais de cette EQM et de ce biais. Un estimateur sans biais de la proportion vraie, $\bar{\mu}$, s'obtient aussi par un traitement de Madow (1965), mais il est impossible dans ce cas d'estimer la variance totale de réponse.

Posons que nous connaissons les valeurs réelles pour un sous-échantillon aléatoire, s', de s. Soit $\bar{x}_{s'}(t)$ l'estimation sans biais tirée du sous-échantillon s' par les valeurs observées à l'essai t, et soit $\bar{\mu}_{s'}$ l'estimation correspondante faisant appel aux valeurs réelles :

$$E_{s'}(\bar{x}_{s'}(t)|s,t) = \bar{x}_s(t)$$

$$E_s E_{s'}(\bar{\mu}_{s'}) = \bar{\mu}.$$

Ainsi, $\hat{B} = (\bar{x}_{s'}(t) - \bar{\mu}_{s'})$ est un estimateur sans biais de B.

Considérons d'abord l'estimateur $(\bar{x}_{s'}(t) - \bar{\mu}_{s'})^2$. Soit E désignant les espérances sur s', s et t (c'est-à-dire, $E = E_t E_s E_{s'}$).

$$E(\bar{x}_{s'}(t) - \bar{\mu}_{s'})^2 = V(\bar{x}_{s'}(t) - \bar{\mu}_{s'}) + [E(\bar{x}_{s'}(t) - \bar{\mu}_{s'})]^2$$

$$= V_t\{E_{ss'}(\bar{x}_{s'}(t) - \bar{\mu}_{s'})|t\} + E_t V_{ss'}\{(\bar{x}_{s'}(t) - \bar{\mu}_{s'})|t\} + B^2.$$

Deuxièmement, soit $v_s(\bar{x}_s(t))$ un estimateur de la variance tel que

$$E_s[v_s(\bar{x}_s(t))|t] = V_s[\bar{x}_s(t)|t],$$

celui-ci étant un estimateur sans biais de la variance d'échantillonnage pour le plan d'échantillonnage de l'enquête et pour un essai donné t.

Troisièmement, soit $v_{ss'}(\bar{x}_{s'}(t) - \bar{\mu}_{s'})$ un estimateur de la variance tel que

$$\underset{ss'}{E}\,[v_{ss'}(\bar{x}_{s'}(t) - \bar{\mu}_{s'})|t] = \underset{ss'}{V}\,[(\bar{x}_{s'}(t) - \bar{\mu}_{s'})|t].$$

Si maintenant nous définissons $\hat{EQM}\,(\bar{x}_s(t))$ par

$$\hat{EQM}(\bar{x}_s(t)) = (\bar{x}_{s'}(t) - \bar{\mu}_{s'})^2 + v_s(\bar{x}_s(t)) - v_{ss'}(\bar{x}_{s'}(t) - \bar{\mu}_{s'}) \quad (2.1)$$

alors

$$E(\hat{EQM}\,(\bar{x}_s(t))) = \underset{t}{V}\,\underset{ss'}{E}\,\{(x_{s'}(t) - \bar{\mu}_{s'})|t\} + \underset{t}{E}\,\underset{ss'}{V}\,\{(\bar{x}_{s'}(t) - \bar{\mu}_{s'})|t\}$$

$$+ B^2 + \underset{t}{E}\,V(\bar{x}_s(t)|t) - \underset{t}{E}\,\underset{ss'}{V}\,\{(\bar{x}_{s'}(t) - \bar{\mu}_{s'})|t\}$$

$$= \underset{t}{V}\,\underset{s}{E}\,(\bar{x}_s(t)|t) + \underset{t}{E}\,\underset{s}{V}\,(\bar{x}_s(t)|t) + B^2$$

$$= V_{t,s}(\bar{x}_s(t)) + B^2$$

$$= EQM\,(\bar{x}_s(t))$$

Dans ce cas, $\hat{EQM}\,(\bar{x}_s(t))$ en (2.1) est un estimateur sans biais de $EQM\,(\bar{x}_s(t))$.

Si l'enquête initiale constitue un recensement, le terme intermédiaire dans $\hat{EQM}\,(\bar{x}_s(t))$ disparaît, et nous avons

$$\hat{EQM}(\bar{X}(t)) = (\bar{x}_{s'}(t) - \bar{\mu}_{s'})^2 - v_{s'}(\bar{x}_{s'}(t) - \bar{\mu}_{s'}) \quad (2.2)$$

vu l'estimateur sans biais, $\hat{B}$, du biais de $\bar{x}_s(t)$, un estimateur sans biais de la proportion vraie, $\bar{\mu}$, s'obtient par

$$\hat{\bar{\mu}} = \bar{x}_s(t) - \hat{B}$$

$$= \bar{x}_s(t) - (\bar{x}_{s'}(t) - \bar{\mu}_{s'}),$$

avec

$$V(\hat{\bar{\mu}}) = \underset{t\,s\,s'}{E\,E\,V}\,(\hat{\bar{\mu}}) + \underset{t\,s\,s'}{E\,V\,E}\,(\hat{\bar{\mu}}) + \underset{t\,s\,s'}{V\,E\,E}\,(\hat{\bar{\mu}})$$

$$= \underset{t\,s\,s'}{E\,E\,V}\,(\bar{x}_{s'}(t) - \bar{\mu}_{s'}) + \underset{t\,s\,s'}{E\,V\,E}\,(\bar{\mu}_{s'})$$

comme $\underset{s'}{E}\,(\hat{\bar{\mu}}|t,s) = \underset{s'}{E}\,(\bar{\mu}_{s'}|t,s)$

et avec $\underset{t\,s\,s'}{V\,E\,E}\,(\hat{\bar{\mu}}) = 0$,

$$\therefore\ V(\hat{\bar{\mu}}) = \underset{t\,s\,s'}{E\,E\,V}\,(\bar{x}_{s'}(t) - \bar{\mu}_{s'}) + \underset{t\,ss'}{E\,V}\,(\bar{\mu}_{s'}) - \underset{t\,s\,s'}{E\,E\,V}\,(\bar{\mu}_{s'}).$$

Si $v_{s'}(\bar{x}_{s'}(t) - \bar{\mu}_{s'})$ est un estimateur de la variance tel que

$$E\{v_{s'}(\bar{x}_{s'}(t) - \bar{\mu}_{s'})|t,s\} = \underset{s'}{V}\,(\bar{x}_{s'}(t) - \bar{\mu}_{s'}),$$

si $v_{ss'}(\bar{\mu}_{s'})$ est un estimateur de la variance tel que

$$E[v_{ss'}(\bar{\mu}_{s'})|t] = \underset{ss'}{V}\,(\bar{\mu}_{s'})$$

et si $v_{s'}(\bar{\mu}_{s'})$ est un estimateur de la variance tel que

$$E[v_{s'}(\bar{\mu}_{s'})|s,t] = \underset{s'}{V}\,(\bar{\mu}_{s'}),$$

un estimateur sans biais de $V(\hat{\bar{\mu}})$ s'obtient par

$$\hat{V}(\hat{\bar{\mu}}) = v_{s'}(\bar{x}_{s'}(t) - \bar{\mu}_{s'}) - v_{s'}(\bar{\mu}_{s'}) + v_{ss'}(\bar{\mu}_{s'}).$$

Si l'enquête constitue un recensement, un estimateur sans biais, $\hat{B}$, du biais de $\bar{X}(t)$ est encore donné par

$$\hat{B} = (\bar{x}_{s'}(t) - \bar{\mu}_{s'})$$

et un estimateur sans biais de la proportion vraie, par

$$\hat{\bar{\mu}} = \bar{X}(t) - (\bar{x}_{s'}(t) - \bar{\mu}_{s'}).$$

Dans l’expression de la variance de $\hat{\bar{\mu}}$ qui vient plus haut du cas de l’enquête par sondage, s devient maintenant la population totale. Les deuxième et troisième termes s’annulent donc, et nous avons

$$V(\hat{\bar{\mu}}) = E\ V\ (\bar{x}_{s'}(t) - \bar{\mu}_{s'}) \ ,$$

et, par conséquent, un estimateur sans biais de $V(\hat{\bar{\mu}})$ s’obtient par

$$V(\hat{\bar{\mu}}) = v_{s'}[\bar{x}_{s'}(t) - \bar{\mu}_{s'}].$$

2.3 Exemple

Dans cette section, nous décrirons comment les estimateurs qui précèdent ont été appliqués à une petite étude réalisée en relation avec le Recensement de 1971, et nous présenterons certains résultats numériques.

Quand les données de ce recensement sur le type de logement sont devenues disponibles, on a soupçonné que certaines catégories, notamment celles des appartements et des duplex, avaient été l’objet d’une sous-déclaration grossière. C’est pourquoi on a lancé une suite de petites études pour reconnaître les sources d’erreur.

Une de ces études a eu lieu dans la région de l’Outaouais. L’objectif était notamment de comparer les réponses des recensés à la question sur le type de logement dans le Recensement de 1971 au type « vrai » de logement selon les observations à vue d’un expert. La comparaison a porté sur tous les ménages de douze secteurs de dénombrement (SD).

Comme cette étude va dans le sens du cadre dressé à la section précédente, nous avons pris les données d’échantillon pour illustrer l’utilisation de ces estimateurs dans une application particulière. Il convient cependant de noter que les chiffres présentés risquent d’accuser une variabilité d’échantillonnage plutôt marquée, puisqu’ils sont fondés sur un très petit échantillon en grappes. Nous décrirons maintenant les estimateurs employés dans le cas qui nous occupe.

Posons que la population totale se divise en K secteurs de dénombrement. Soit M_k la taille du $k^{ième}$ SD et $\bar{M}$ la taille moyenne des SD. Posons ensuite qu’un échantillon aléatoire simple de k_o SD est prélevé sur le total, K, et que, dans chaque SD, toutes les unités sont observées.

Désignons cet échantillon par s' et supposons que les valeurs réelles sont déterminées pour chaque unité constitutive.

Définissons

$x_{ik}(t)$: valeur observée de l'unité dans SD k à l'essai t,

μ_{ik}: valeur réelle de l'unité dans SD k,

$e_{ik}(t) = x_{ik}(t) - \mu_{ik}$: écart de réponse de l'unité i dans SD k à l'essai t.

Dans ce cas,

$$\bar{x}_{s'}(t)^* = \frac{1}{k_o \bar{M}} \sum_{k \in s'} \sum_{i} x_{ik}(t), \quad \bar{\mu}_{s'} = \frac{1}{k_o \bar{M}} \sum_{k \in s'} \sum_{i} \mu_{ik}$$

$$\bar{e}(t) = (\bar{x}_{s'}(t) - \bar{\mu}_{s'}).$$

De même,

$$\bar{x}_k(t) = \frac{1}{M_k} \sum_{i=1}^{M_k} x_{ik}(t), \quad \bar{\mu}_k = \frac{1}{M_k} \sum_{i=1}^{M_k} \mu_{ik}.$$

Un estimateur sans biais de la variance d'échantillonnage $v_{s'}[\bar{x}_{s'}(t) - \bar{\mu}_{s'}]$ pour ce plan d'échantillonnage s'obtient par

$$v_{s'}(\bar{x}_{s'}(t) - \bar{\mu}_{s'}) = v_{s'}(\bar{e}(t))$$

$$= \frac{K-k_o}{K-k_o} \cdot \frac{1}{k_o - 1} \sum_{k=1}^{k_o} \left[\sum_{i=1}^{M_k} \frac{e_{ik}(t)}{\bar{M}} - \bar{e}(t) \right]^2$$

$$= \frac{K-k_o}{K(k_o-1)} \cdot \frac{1}{k_o \bar{M}^2} \left\{ \sum_{k=1}^{k_o} \left(\sum_{i=1}^{M_k} e_{ik}(t) \right)^2 - k_o \bar{M}^2 \bar{e}^2(t) \right\} .$$

Si on insère cette expression dans la formule EQM à l'équation 2, on dégage l'estimateur sans biais suivant de EQM $[\bar{x}(t)]$.

*Nous avons employé cet estimateur pour les raisons suivantes :
1. il est exempt de biais et correspond donc à $\bar{x}_{s'}(t)$ dans la théorie de l'estimation;
2. les SD ne varient pas en taille outre mesure

$$\hat{EQM}[\bar{x}(t)] = \frac{1}{K(k_o-1)} \{k_o(K-1)(\bar{x}_{s'}(t)-\bar{\mu}_{s'})^2 - \frac{(K-k_o)}{\bar{M}^2 k_o} \sum_{k=1}^{k_o} M_k^2(\bar{x}_k(t)-\bar{\mu}_k)^2\} \quad (2.3)$$

si, comme c'est souvent le cas, $\bar{X}(t)$ représente la proportion de recensement de la population totale dans une catégorie quelconque, tant $x_{ik}(t)$ que μ_{ik} sont des variables 0-1. L'échantillon de SD k peut donc se répartir selon le tableau suivant des fréquences.

Tableau 1
Classement de recensement $x_{ik}(t)$

Classement « vrai » μ_{ik}	1	0	TOTAL	
1	a_k	b_k	a_k+b_k	
0	c_k	d_k	c_k+d_k	
TOTAL	a_k+c_k	b_k+d_k	M_k	

Un terme servant fréquemment à mesurer les erreurs de classement au recensement est le taux net de différence, qui se définit de la manière suivante pour SD k :

$$r_k = \frac{c_k-b_k}{M_k} .$$

Pour l'échantillon total, nous définissons ce taux r comme

$$r = \frac{c-b}{k_o\bar{M}} \quad \text{où} \quad c = \sum_{k=1}^{k_o} c_k, \quad b = \sum_{k=1}^{k_o} b_k .$$

La quantité r est identique à $\bar{e}(t)$ et livre donc un estimateur sans biais du biais de la statistique de recensement, $\bar{X}(t)$.

Dans ses termes r et r_k, l'équation (3) peut ainsi s'exprimer :

$$\hat{EQM}\ [x(t)] = \frac{1}{K(k_o-1)} \{k_o(K-1)\ r^2 - \frac{(K-k_o)}{\bar{M}^2 k_o} \sum_{k=1}^{k_o} M_k^2\ r_k^2\}.$$

Si nous posons que k_o est grand et que $k_o << K$, il est possible de simplifier davantage cette expression sous la forme suivante :

$$\hat{EQM}\ [\bar{x}(t)] = r^2 - \frac{1}{(\bar{M}\ k_o)^2} \sum_{k=1}^{k_o} M_k^2\ r_k^2 .$$

De même, on pourra exprimer la variance d'échantillonnage $v_{s'}\ (\bar{x}_{s'}(t) - \bar{\mu}_{s'})$ dans les termes r et r_k sous la forme simplifiée suivante :

$$v_{s'}(\bar{x}_{s'}(t) - \bar{\mu}_{s'}) = \frac{1}{(\bar{M}\ k_o)^2} \sum_{k=1}^{k_o} M_k^2\ r_k^2 - \frac{r^2}{k_o} .$$

Ces deux expressions peuvent facilement se calculer à partir du tableau des fréquences plus haut.

Le tableau 2 résume les résultats numériques qui confirment l'hypothèse initiale selon laquelle les appartements et les duplex ont été sous-estimés au Recensement de 1971.

Tableau 2
Mesures de fiabilité des statistiques du recensement sur le type de logement

Type de logement	Pourcentage au recensement[1] $\bar{x}(t)\times 100\%$	REQM $(\bar{x}(t))$	Biais $(\bar{x}(t))$	Erreur-type du biais estimé $(\bar{x}(t))$	Pourcentage vrai estimé	Erreur-type[2] du pourcentage vrai estimé
	%	%	%	%	%	%
Logement individuel	46,4	1,38	1,69	0,98	44,7	0,98
Logement double	7,6	2,08	2,20	0,70	5,4	0,70
Duplex	8,1	2,06	-2,36	1,17	10,5	1,17
Logement unifamilial attenant ou en rangée	8,3	0,44	0,62	0,45	7,7	0,45
Appartement	29,2	2,26	-2,42	0,88	31,6	0,88

1. La population totale comprend 0,4 % de maisons mobiles.
2. Cela équivaut à l'erreur-type du biais.

2 Approche II

Dans cette approche, nous employons comme mesure de l'erreur l'écart entre un chiffre particulier du recensement et le paramètre vrai estimé.

Le traitement est différent de celui du modèle habituel de l’erreur de réponse dans la mesure où aucun modèle de probabilités n’est posé concernant les observations faites au recensement.

Soit $\bar{\mu}$ la moyenne de population estimée et $\bar{x}$, le chiffre de recensement correspondant. Dans ce cas, l’erreur nette tenant à l’utilisation de $\bar{x}$ comme estimation de $\bar{\mu}$ s’obtient par

$$E = \bar{x} - \bar{\mu}.$$

Posons que, pour un échantillon s' de la population, les valeurs réelles peuvent être déterminées. Soit $\bar{x}_{s'}$ et $\bar{\mu}_{s'}$ des estimations sans biais de $\bar{x}$ et $\bar{\mu}$ respectivement tirées de l’échantillon. Un estimateur sans biais de E s’obtient par

$$\hat{E} = \bar{x}_{s'} - \bar{\mu}_{s'}.$$

On peut employer l’estimation qui précède pour produire un nouvel estimateur de $\bar{\mu}$ en corrigeant $\bar{x}$ de la manière suivante :

$$\hat{\bar{\mu}} = \bar{x} - (\bar{x}_{s'} - \bar{\mu}_{s'}),$$

expression pour laquelle la variance d’échantillonnage s’obtient par

$$V(\hat{\bar{\mu}}) = V(\bar{x}_{s'} - \bar{\mu}_{s'}).$$

On peut aisément démontrer que $\hat{\bar{\mu}}$ sera plus fiable que $\bar{\mu}_s$ quand $\bar{x}_{s'}$ et $\bar{\mu}_{s'}$ sont en étroite corrélation, ce qui est normalement le cas.

Cette constatation pourrait aussi valoir pour les enquêtes par sondage où les valeurs réelles sont connues pour un sous-échantillon de l’échantillon de départ.

4 Conclusion

Nous avons exposé deux modes de mesure de la fiabilité des données de recensement et d’enquête par sondage lorsque les valeurs réelles peuvent être déterminées pour un échantillon ou un sous-échantillon de la

population. Pour chaque approche décrite, nous avons en outre présenté une méthode de correction de l'estimation initiale.

Cette théorie a été conçue surtout pour les applications destinées aux problèmes d'erreur de réponse au recensement, mais elle a aussi sa place dans le traitement d'autres types de cas : problèmes d'erreur de dénombrement, de codage, etc. L'approche à retenir dépendrait donc en général de la nature des erreurs étudiées.

Bibliographie

Madow (1965). On some aspects of response error measurement. *Proceedings of the ASA*, Social Statistics Section, 182-192.