Techniques d'enquête, juin 1994
Vol. 20, n° 1, pp. 57–61
Statistique Canada

Estimation pour grappes chevauchantes lorsque la taille de la population est inconnue

D.S. TRACY et S.S. OSAHAN[1]

RÉSUMÉ

Singh (1988) propose deux méthodes d'échantillonnage en vue d'estimer la moyenne d'une population en grappes chevauchantes lorsque la taille de la population est connue. Dans cet article, nous étudions des estimateurs par quotient appliqués dans ces deux méthodes en supposant que la taille réelle de la population est inconnue, ce qui est plus conforme à la réalité des enquêtes par sondage. Nous comparons l'efficacité des estimateurs appliqués dans l'une et l'autre méthodes et nous donnons un exemple numérique.

MOTS CLÉS: Grappes chevauchantes; formation de grappes avant échantillonnage; erreur quadratique moyenne; efficacité relative.

1. INTRODUCTION

Dans l'échantillonnage par grappes, les grappes sont formées soit avant l'échantillonnage (FGAVÉ), soit après l'échantillonnage (FGAPÉ). Dans les deux cas, les grappes peuvent se chevaucher ou non. Il existe déjà beaucoup d'ouvrages de recherche sur les grappes non chevauchantes. Cependant, de nombreux cas d'échantillonnage impliquent des grappes chevauchantes. Par exemple, on peut retrouver de telles grappes dans une enquête épidémiologique régionale portant sur une maladie contagieuse comme la tuberculose, qui tend à se répandre avec la propagation du SIDA (Gifford-Jones 1993). Dans ce cas-ci, les grappes peuvent être formées de sujets infectés ou de personnes qui sont en rapport étroit avec eux et qui sont plus vulnérables à cette maladie. De même, dans une étude écologique, les grappes peuvent être formées de centrales au charbon qui émettent des hydrocarbures aromatiques polycycliques (HAP), composés doués de propriétés cancérogènes. Les grappes sont constituées en fonction de la concentration de ces gaz, et des études peuvent s'imposer dans le but de surveiller la pollution atmosphérique, qui est à l'origine d'infections pulmonaires telles la bronchite. Au sujet des grappes chevauchantes, on peut consulter les travaux effectués par Goel et Singh (1977), Agarwal et Singh (1982) et Amdekar (1985) sur certains aspects de la question. Cependant, les méthodes élaborées par ces auteurs présentent toutes des lacunes.

Il y a quelques années, Singh (1988) a élaboré un estimateur très simple pour la moyenne d'une population. Il utilise cet estimateur dans deux méthodes d'échantillonnage par grappes (selon la formule FGAVÉ) en supposant que la taille de la population est connue. La première méthode prévoit un échantillonnage des grappes avec probabilités égales tandis que la seconde prévoit un échantillonnage avec probabilités proportionnelles à la taille de la grappe. En ce qui concerne l'échantillonnage des éléments des grappes, la règle des probabilités égales s'applique pour les deux méthodes. Toutefois, supposer que la taille de la population est connue n'est pas réaliste. Si c'était vraiment le cas, on connaîtrait *a priori* tous les éléments en double de la population et il suffirait de les éliminer pour accroître l'efficacité du plan d'échantillonnage. C'est pourquoi il est nécessaire d'améliorer les estimateurs de la moyenne d'une population de Singh (1988) afin de les rendre plus utiles, car ils dépendent de la taille réelle de la population. C'est exactement le but du présent article.

Nous proposons deux méthodes d'échantillonnage par grappes (formule FGAVÉ) qui utilisent des estimateurs par quotient ordinaires de la moyenne de population; ces estimateurs ne dépendent pas de la taille réelle de la population. Comme dans Singh (1988), la première méthode consiste en un échantillonnage avec remise avec probabilités égales, tandis que la seconde consiste en un échantillonnage avec probabilités inégales. Dans les deux cas, les éléments des grappes sont prélevés suivant un plan d'échantillonnage sans remise avec probabilités égales.

La population de N unités à l'étude peut être définie comme un ensemble de K grappes chevauchantes où N_i désigne le nombre d'unités contenues dans la grappe i et $\sum_{i=1}^{K} N_i = M \geq N$, la taille réelle (inconnue) de la population, (l'égalité ne vaut que dans le cas des grappes non chevauchantes). Une unité peut appartenir à plus d'une grappe à la fois. Posons y comme la caractéristique étudiée et $\bar{Y}$, comme la moyenne de la population.

Définissons

$$Z_{ij} = Y_{ij}/F_{ij}, \quad W_{ij} = 1/F_{ij}; \quad i = 1, 2, \ldots, K, \quad \text{et} \quad j = 1, 2, \ldots, N_i$$

où Y_{ij} est la valeur de y pour l'unité j dans la grappe i et F_{ij} est le nombre de fois que cette unité est observée dans les K grappes.

[1] D.S. Tracy et S.S. Osahan, Department of Mathematics and Statistics, University of Windsor, Windsor (Ontario), N9B 3P4.

Lorsqu'on dispose de données informatiques par grappe sur les unités, il est facile de connaître la valeur des F_{ij}. En reprenant l'exemple en épidémiologie cité plus haut, supposons qu'il existe des données informatiques sur des ménages ou des particuliers, accompagnées de codes d'identification pour ces unités (par ex., numéro civique, numéro d'assurance sociale ou numéro d'assurance-maladie). Alors, par une simple instruction de programme, un expérimentateur peut savoir facilement, grâce au code d'identification, combien de fois une certaine unité se répète dans les différentes grappes. En outre, si nous avons un diagramme des grappes chevauchantes et si le critère de formation des grappes ne permet pas d'éliminer les doubles, il est possible de connaître le nombre de fois que des unités sont observées dans les différentes grappes.

Nous examinons les deux méthodes d'échantillonnage dans la section 2 et nous comparons leur efficacité dans la section 3.

2. DESCRIPTION DES DEUX MÉTHODES

Les deux méthodes proposées sont décrites dans les sections 2.1 et 2.2. Nous en faisons la comparaison dans la section 3.

2.1 Méthode A

Cette méthode comporte les étapes suivantes:

a) Prélever k grappes parmi K au moyen d'un échantillonnage aléatoire simple avec remise (ÉASAR).

b) Dans la grappe échantillonnée i de taille $N_i (i = 1, \ldots, K)$, prélever n_i unités élémentaires par échantillonnage aléatoire simple sans remise (ÉASSR).

Théorème 1. L'estimateur par quotient suivant un ÉAS

$$\bar{z}_{RS} = \hat{Y}_{RS}/\hat{N}_{RS} = \frac{K}{k}\sum_{i=1}^{k} N_i\bar{z}_i \Big/ \frac{K}{k}\sum_{i=1}^{k} N_i\bar{w}_i \tag{1}$$

a pour biais relatif (au premier degré d'approximation)

$$BR(\bar{z}_{RS}) \doteq \frac{K}{k}\left[\left(\frac{\sigma_{bw}^2}{N^2} - \frac{\sigma_{bzw}}{NY}\right)K + \sum_{i=1}^{K} N_i^2\left(\frac{1}{n_i} - \frac{1}{N_i}\right)\left(\frac{S_{iw}^2}{N^2} - \frac{S_{izw}}{NY}\right)\right] \tag{2}$$

où

$$\sigma_{bzw} = \sum_{i=1}^{K} (N_i\bar{Z}_i - Y/K)(N_i\bar{W}_i - N/K)/K$$

$$S_{izw} = \sum_{j=1}^{N_i} (Z_{ij} - \bar{Z}_i)(W_{ij} - \bar{W}_i)/(N_i - 1),$$

$$\bar{Z}_i = \sum_{j=1}^{N_i} Z_{ij}/N_i \quad \text{et} \quad \bar{z}_i = \sum_{j=1}^{n_i} z_{ij}/n_i,$$

et σ_{bw}^2, S_{iw}^2, $\bar{W}_i$ et $\bar{w}_i$ correspondent respectivement à σ_{bzw}, S_{izw}, $\bar{Z}_i$ et $\bar{z}_i$, z étant remplacé par w et Y, par N.

Démonstration. D'après un résultat courant, le biais relatif de l'estimateur $\bar{z}_{RS}$ (au premier degré d'approximation) est

$$BR(\bar{z}_{RS}) \doteq [V(\hat{N}_{RS})/N^2] - \mathrm{Cov}(\hat{Y}_{RS},\hat{N}_{RS})/YN. \tag{3}$$

Posons E_2 et V_2 comme l'espérance et la variance conditionnelles relatives à un échantillon donné de grappes, et E_1 et V_1 comme l'espérance et la variance relatives à tous les échantillons de grappes. Nous avons alors

$$\begin{aligned} V(\hat{N}_{RS}) &= V_1E_2(\hat{N}_{RS}) + E_1V_2(\hat{N}_{RS}) \\ &= V_1\left[\frac{K}{k}\sum_{i=1}^{k} N_iE_2(\bar{w}_i)\right] + E_1\left[\frac{K^2}{k^2}\sum_{i=1}^{k} N_i^2V_2(\bar{w}_i)\right] \\ &= V_1\left(\frac{K}{k}\sum_{i=1}^{k} N_i\bar{W}_i\right) + E_1\left[\frac{K^2}{k^2}\sum_{i=1}^{k} N_i^2\left(\frac{1}{n_i} - \frac{1}{N_i}\right)S_{iw}^2\right] \\ &= \frac{K^2}{k}\sigma_{bw}^2 + \frac{K}{k}\sum_{i=1}^{K} N_i^2\left(\frac{1}{n_i} - \frac{1}{N_i}\right)S_{iw}^2. \end{aligned} \tag{4}$$

De la même manière, nous avons

$$\mathrm{Cov}(\hat{Y}_{RS},\hat{N}_{RS}) = \frac{K^2}{k}\sigma_{bzw} + \frac{K}{k}\sum_{i=1}^{K} N_i^2\left(\frac{1}{n_i} - \frac{1}{N_i}\right)S_{izw}. \tag{5}$$

En substituant (4) et (5) dans (3), nous obtenons l'équation (2), ce qui complète la démonstration du théorème.

Théorème 2. L'erreur quadratique moyenne (EQM) de l'estimateur $\bar{z}_{RS}$ (au premier degré d'approximation) est

$$\mathrm{EQM}(\bar{z}_{RS}) \doteq$$

$$\frac{K}{kN^2}\sum_{i=1}^{K} N_i^2\left[(\bar{Z}_i - \bar{Y}\bar{W}_i)^2 + \left(\frac{1}{n_i} - \frac{1}{N_i}\right)D_i^2\right] \qquad (6)$$

où $D_i^2 = S_{iz}^2 - 2\bar{Y}S_{izw} + \bar{Y}^2 S_{iw}^2$, et $S_{iz}^2 = \sum_{j=i}^{N_i}(Z_{ij} - \bar{Z}_i)^2/(N_i - 1)$.

Démonstration. Au premier degré d'approximation, nous avons

$$\mathrm{EQM}(\bar{z}_{RS}) \doteq [V(\hat{Y}_{RS}) - 2\bar{Y}\,\mathrm{Cov}(\hat{Y}_{RS}, \hat{N}_{RS}) + \bar{Y}^2 V(\hat{N}_{RS})]/N^2. \qquad (7)$$

D'après (4), on peut exprimer $V(\hat{Y}_{RS})$ par la formule

$$V(\hat{Y}_{RS}) = \frac{K^2}{k}\sigma_{bz}^2 + \frac{K}{k}\sum_{i=1}^{K} N_i^2\left(\frac{1}{n_i} - \frac{1}{N_i}\right)S_{iz}^2 \qquad (8)$$

où $\sigma_{bz}^2 = \sum_{i=1}^{K} (N_i\bar{Z}_i - Y/K)^2/K$.

En substituant (4), (5) et (8) dans l'équation (7), nous obtenons, après simplification,

$$\mathrm{EQM}(\bar{z}_{RS}) \doteq \frac{K^2}{kN^2}(\sigma_{bz}^2 - 2\bar{Y}\sigma_{bzw} + \bar{Y}^2\sigma_{bw}^2)$$

$$+ \frac{K}{kN^2}\sum_{i=1}^{K} N_i^2\left(\frac{1}{n_i} - \frac{1}{N_i}\right)(S_{iz}^2 - 2\bar{Y}S_{izw} + \bar{Y}^2 S_{iw}^2). \qquad (9)$$

En substituant dans (9) les expressions pour σ_{bz}^2, σ_{bzw} et σ_{bw}^2, et en simplifiant, nous obtenons l'équation (6). Dans le théorème suivant, nous définissons un estimateur de $\mathrm{EQM}(\bar{z}_{RS})$.

Théorème 3. Un estimateur convergent de $\mathrm{EQM}(\bar{z}_{RS})$ (au premier degré d'approximation) est défini

$$\widehat{\mathrm{EQM}}(\bar{z}_{RS}) = \frac{K^2}{k\hat{N}_{RS}^2}\cdot\frac{1}{k-1}\sum_{i=1}^{k} N_i^2(\bar{z}_i - \bar{z}_{RS}\bar{w}_i)^2. \qquad (10)$$

Démonstration. Nous remarquons que le premier degré d'échantillonnage consiste en un ÉASAR et que les variables aléatoires $N_i\bar{z}_i$ et $N_i w_i$, contenues dans l'estimateur par quotient, sont distribuées de façon indépendante mais identique. On peut donc estimer l'erreur quadratique moyenne de $\bar{z}_{RS}$ en se fondant sur le résultat bien connu selon lequel l'estimateur de la variance dans un plan à plusieurs degrés peut ne prendre en compte que le premier degré (voir Särndal, Swensson et Wretman, 1992, résultats 2.9.1 et 4.5.1).

D'après (9), un estimateur sans biais de

$$\sigma_{bz}^2 + \frac{1}{K}\sum_{i=1}^{K} N_i^2\left(\frac{1}{n_i} - \frac{1}{N_i}\right)S_{iz}^2$$

peut être défini

$$s_{bz}^2 = \frac{1}{k-1}\sum_{i=1}^{k}\left(N_i\bar{z}_i - \sum_{i=1}^{k} N_i\bar{z}_i/k\right)^2, \qquad (11)$$

et un estimateur sans biais de

$$\sigma_{bzw} + \frac{1}{K}\sum_{i=1}^{K} N_i^2\left(\frac{1}{n_i} - \frac{1}{N_i}\right)S_{izw}$$

est

$$s_{bzw} = \frac{1}{k-1}\sum_{i=1}^{k}\left(N_i\bar{z}_i - \sum_{i=1}^{k} N_i\bar{z}_i/k\right) \times \left(N_i\bar{w}_i - \sum_{i=1}^{k} N_i\bar{w}_i/k\right). \qquad (12)$$

De la même manière, un estimateur indépendant de

$$\sigma_{bw}^2 + \frac{1}{K}\sum_{i=1}^{K} N_i^2\left(\frac{1}{n_i} - \frac{1}{N_i}\right)S_{iw}^2$$

est s_{bw}^2, défini comme en (11).

À l'aide de ces résultats, il est facile de montrer qu'un estimateur convergent de $\mathrm{EQM}(\bar{z}_{RS})$ (équ. 6) est défini par l'expression

$$\widehat{\mathrm{EQM}}(\bar{z}_{RS}) = \frac{K^2}{k\hat{N}_{RS}^2}(s_{bz}^2 - 2\bar{z}_{RS}s_{bzw} + \bar{z}_{RS}^2 s_{bw}^2),$$

qui peut s'écrire comme en (10).

2.2 Méthode B

Cette méthode comporte les étapes suivantes:

a) Prélever k grappes parmi K au moyen d'un échantillonnage avec remise avec probabilité proportionnelle à la taille (ÉARPPT), les probabilités de sélection étant $P_i = N_i/M$, $i = 1, \ldots, K$.

b) Même étape que celle de la méthode A.

Théorème 4. L'estimateur par quotient suivant un échantillonnage PPT,

$$\bar{z}_{RP} = \hat{Y}_{RP}/\hat{N}_{RP} = \frac{M}{k}\sum_{i=1}^{k}\bar{z}_i \Big/ \frac{M}{k}\sum_{i=1}^{k}\bar{w}_i \qquad (13)$$

a pour biais relatif (au premier degré d'approximation)

$$BR(\bar{z}_{RP}) \doteq \frac{M^2}{k}\left[\left(\frac{\sigma^2_{bw'}}{N^2} - \frac{\sigma_{bzw'}}{YN}\right) + \sum_{i=1}^{K} \frac{N_i}{M}\left(\frac{1}{n_i} - \frac{1}{N_i}\right)\left(\frac{S^2_{iw}}{N^2} - \frac{S_{izw}}{YN}\right)\right] \quad (14)$$

où

$$\sigma_{bzw'} = \sum_{i=1}^{K} (\bar{Z}_i - Y/M)(\bar{W}_i - N/M)(N_i/M)$$

et $\sigma^2_{bw'}$ correspond à $\sigma_{bzw'}$, z étant remplacé par w et Y, par N.

Démonstration. D'après un résultat courant, le biais relatif approximatif (au premier degré d'approximation) est

$$BR(\bar{z}_{RP}) \doteq [V(\hat{N}_{RP})/N^2] - \mathrm{Cov}(\hat{Y}_{RP},\hat{N}_{RP})/YN. \quad (15)$$

Nous avons

$$V(\hat{N}_{RP}) = V_1E_2(\hat{N}_{RP}) + E_1V_2(\hat{N}_{RP})$$

$$= M^2\left[V_1\frac{1}{k}\sum_{i=1}^{k} E_2(\bar{w}_i) + E_1\frac{1}{k^2}\sum_{i=1}^{k} V_2(\bar{w}_i)\right]$$

$$= M^2\left[V_1\frac{1}{k}\sum_{i=1}^{k} \bar{W}_i + E_1\frac{1}{k^2}\sum_{i=1}^{k} \left(\frac{1}{n_i} - \frac{1}{N_i}\right)S^2_{iw}\right]$$

$$= \frac{M^2}{k}\left[\sigma^2_{bw'} + \sum_{i=1}^{K} \frac{N_i}{M}\left(\frac{1}{n_i} - \frac{1}{N_i}\right)S^2_{iw}\right]. \quad (16)$$

De la même manière, nous pouvons écrire

$$\mathrm{Cov}(\hat{Y}_{RP},\hat{N}_{RP}) = \frac{M^2}{k}\left[\sigma_{bzw'} + \sum_{i=1}^{K} \frac{N_i}{M}\left(\frac{1}{n_i} - \frac{1}{N_i}\right)S_{izw}\right]. \quad (17)$$

En substituant (16) et (17) dans (15), nous obtenons l'équation (14).

Théorème 5. L'EQM de l'estimateur $\bar{z}_{RP}$ (au premier degré d'approximation) est définie

$$\mathrm{EQM}(\bar{z}_{RP}) \doteq \frac{M}{kN^2}\sum_{i=1}^{K} N_i \times \left[(\bar{Z}_i - \bar{Y}\bar{W}_i)^2 + \left(\frac{1}{n_i} - \frac{1}{N_i}\right)D_i^2\right]. \quad (18)$$

Démonstration. Au premier degré d'approximation, nous écrivons

$$\mathrm{EQM}(\bar{z}_{RP}) \doteq [V(\hat{Y}_{RP}) - 2\bar{Y}\,\mathrm{Cov}(\hat{Y}_{RP},\hat{N}_{RP}) + \bar{Y}^2V(\hat{N}_{RP})]/N^2. \quad (19)$$

En outre, d'après le théorème 2.5 de Singh (1988), nous avons, par analogie,

$$V(\hat{Y}_{RP}) = \frac{M^2}{k}\sigma^2_{bz'} + \frac{M^2}{k}\sum_{i=1}^{K}\frac{N_i}{M}\left(\frac{1}{n_i} - \frac{1}{N_i}\right)S^2_{iz}, \quad (20)$$

où $\sigma^2_{bz'} = \sum_{i=1}^{K} (N_i/M)(\bar{Z}_i - Y/M)^2$. En substituant (16), (17) et (20) dans l'équation (19), puis en simplifiant, nous obtenons l'équation (18).

Théorème 6. Un estimateur convergent de $\mathrm{EQM}(\bar{z}_{RP})$ (au premier degré d'approximation) est défini

$$\widehat{\mathrm{EQM}}(\bar{z}_{RP}) = \frac{M^2}{\hat{N}^2_{RP}}\cdot\frac{1}{k(k-1)}\sum_{i=1}^{k}(\bar{z}_i - \bar{z}_{RP}\bar{w}_i)^2. \quad (21)$$

Démonstration. Comme les unités de sondage du premier degré sont prélevées suivant un ÉARPPT, l'argument présenté dans la démonstration du théorème 3 vaut pour cette démonstration-ci.

D'après l'équation (20) et les résultats 2.9.1 et 4.5.1 de Särndal, Swensson et Wretman (1992), nous pouvons définir un estimateur non biaisé de

$$\sigma^2_{bz'} + \sum_{i=1}^{K}\frac{N_i}{M}\left(\frac{1}{n_i} - \frac{1}{N_i}\right)S^2_{iz}$$

par l'expression

$$s^2_{bz'} = \frac{1}{k-1}\sum_{i=1}^{k}\left(\bar{z}_i - \sum_{i=1}^{k}\bar{z}_i/k\right)^2. \quad (22)$$

De la même manière, si nous définissons $s_{bzw'}$ et $s^2_{bw'}$, nous pouvons montrer que

$$\widehat{\mathrm{EQM}}(\bar{z}_{RP}) = \frac{M^2}{\hat{N}^2_{RP}k}(s^2_{bz'} - 2\bar{z}_{RP}s_{bzw'} + \bar{z}^2_{RP}s^2_{bw'}),$$

qui peut s'écrire comme en (21).

3. COMPARAISON DE L'EFFICACITÉ

Dans cette section, nous comparons l'efficacité des estimateurs utilisés respectivement dans les méthodes A et B.

Remarque. L'estimateur $\bar{z}_{RS}$, de la méthode B, devrait normalement être plus efficace que l'estimateur $\bar{z}_{RP}$, de la méthode A. Voici pourquoi. D'après (6) et (18), nous avons

Voici pourquoi. D'après (6) et (18), nous avons

$$\mathrm{EQM}(\bar{z}_{RS}) - \mathrm{EQM}(\bar{z}_{RP}) \doteq \frac{M}{kN^2} \sum_{i=1}^{K} N_i$$

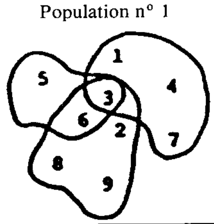
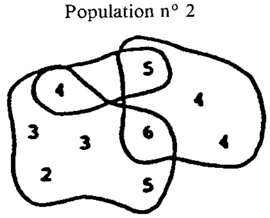
$$\times \left[(\bar{Z}_i - \bar{Y}\bar{W}_i)^2 + \left(\frac{1}{n_i} - \frac{1}{N_i}\right) D_i^2 \right] \left(\frac{KN_i}{M} - 1\right).$$

Si la taille de la grappe, N_i, augmente, le facteur $(KN_i/M - 1)$ augmentera aussi. L'autre facteur du terme en sommation est $N_i[(\bar{Z}_i - \bar{Y}\bar{W}_i)^2 + (1/n_i - 1/N_i)D_i^2]$, il représente la part de l'EQM de $\bar{z}_{RP}$ (équ. 18) attribuable à la variabilité de z et de w dans la grappe i (sans la constante M/kN^2). Si la taille de la grappe, N_i, augmente, la part de EQM $(\bar{z}_{RP})$ attribuable à la grappe i devrait aussi augmenter, ce qui fait que la covariance des deux facteurs est positive. Donc, l'estimateur $\bar{z}_{RP}$ devrait avoir une EQM moins élevée que celle de $\bar{z}_{RS}$.

Exemple numérique. Dans cet exemple, nous appliquons les deux méthodes d'échantillonnage proposées à deux petites populations afin de mettre en lumière le calcul des valeurs F_{ij}, Z_{ij} et W_{ij} et la comparaison de ces valeurs. Pour les deux populations, $K = 3$, $k = 2$, $M = 12$ et $N = 9$. Il y a chevauchement dès qu'une unité est présente dans plus d'une grappe à la fois. Les populations sont décrites dans le tableau 1.

Tableau 1

Comparaison des deux méthodes pour deux petites populations

	Population n° 1			Population n° 2		
						
N_i	3	4	5	2	4	6
n_i	1	2	2	1	2	2
Y_{ij}	3,5,6	1,3,4,7	2,3,6,8,9	4,5	4,4,5,6	2,3,3,4,5,6
F_{ij}	3,1,2	1,3,1,1	1,3,2,1,1	2,2	1,1,2,2	1,1,1,2,1,2
Z_{ij}	1,5,3	1,1,4,7	2,1,3,8,9	2,2.5	4,4,2.5,3	2,3,3,2,5,3
W_{ij}	⅓,1,½	1,⅓,1,1	1,⅓,½,1,1	½,½	1,1,½,½	1,1,1,½,1,½
F	1.38	10.16	18.12	.24	.77	2.94
EQM $(\bar{z}_{RS})$		2.09			0.45	
EQM $(\bar{z}_{RP})$		1.83			0.33	
E.R.		114.21			136.36	
B.R. $(\bar{z}_{RS})$		−.0105			.0348	
B.R. $(\bar{z}_{RP})$		−.0047			−.0037	

L'analyse des résultats du tableau 1 confirme la théorie exposée dans cet article. Dans le cas des deux populations, le facteur $F = N_i[(\bar{Z}_i - \bar{Y}\bar{W}_i)^2 + (1/n_i - 1/N_i)D_i^2]$ augmente avec N_i, ce qui fait que EQM $(\bar{y}_{RP}) <$ EQM $(\bar{y}_{RS})$, comme nous le disions plus haut.

4. CONCLUSION

Dans cet article, nous avons pu examiner la question du chevauchement des grappes sans nous embarrasser de l'hypothèse de la taille de population connue posée par Singh (1988). En outre, nous avons pu comparer les deux méthodes de façon plus directe, alors que Singh devait s'appuyer sur les résultats de Hansen et Hurwitz (1943).

REMERCIEMENTS

Cette étude a été rendue possible en partie grâce à une subvention (n° A-3111) du Conseil de recherches en sciences naturelles et en génie du Canada. Nous tenons aussi à exprimer toute notre reconnaissance aux arbitres et au rédacteur en chef, qui, par les précieux commentaires qu'ils ont faits sur une version antérieure de l'article, ont contribué à améliorer le produit.

BIBLIOGRAPHIE

AGARWAL, D.K., et SINGH, P. (1982). On cluster sampling strategies using ancillary information. *Sankhyā*, B, 44, 184-192.

AMDEKAR, S.J.(1985). An unbiased estimator in overlapping clusters. *Bulletin of the Calcutta Statistical Association*, 15, 231-232.

GIFFARD-JONES, W. (1993). The doctor game. *The Windsor Star*, April 15, 1993.

GOEL, B.B.P.S., et SINGH, D. (1977). On the formation of clusters. *Journal of the Indian Society of Agricultural Statistics*, 29, 53-68.

HANSEN, M.H., et HURWITZ, W.N. (1943). On the theory of sampling from finite populations. *Annals of Mathematical Statistics*, 14, 333-362.

SÄRNDAL, C-E., SWENSSON, B., et WRETMAN, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

SINGH, S. (1988). Estimation in overlapping clusters. *Communications in Statistics, Theory and Methods*, 17, 613-621.