Examen de sondage

Tous les documents sont autorisés. Les exercices sont indépendants.

PARTIE I

1. Rappeler l'estimateur non biaisé du total d'une population, dans le cas du tirage d'un nombre n fixé d'unités avec probabilités inégales et sans remise (PISR). Indiquer avec précision à quoi correspond chaque quantité dans la formule. On le notera $\widehat{Y}_1$.
2. Rappeler la variance de cet estimateur.
3. Montrer que cette variance peut aussi s'écrire ainsi
$$\mathrm{Var}\left[\widehat{Y}_1\right] = \sum_{j=1}^{N}\sum_{k>j}^{N}(\pi_j\pi_k - \pi_{jk})\left(\frac{Y_j}{\pi_j} - \frac{Y_k}{\pi_k}\right)^2,$$
où π_{jk} est la probabilité que les unités j et k fassent simultanément partie de l'échantillon sélectionné.
4. Montrer que cette variance peut encore s'écrire
$$\mathrm{Var}\left[\widehat{Y}_1\right] = \sum_{j=1}^{N}\frac{1-\pi_j}{\pi_j}Y_j^2 + 2\sum_{j=1}^{N}\sum_{k>j}^{N}\frac{(\pi_{jk} - \pi_j\pi_k)}{\pi_j\pi_k}Y_jY_k.$$
5. À partir de cette dernière relation, donner un estimateur de la variance du total. On rappelle que cet estimateur a été proposé par HORVITZ et THOMPSON en 1952.
6. Rappeler la formule de l'estimateur de la variance du total que vous avez dans vos notes de cours. On rappelle que celle-ci a été proposée par YATES et GRUNDY en 1953.
7. Montrer que ces deux estimateurs de la variance du total sont non biaisés.

PARTIE II

Ces deux estimateurs sont très instables et peuvent conduire à des valeurs négatives. GOURIEROUX en 1981, signale que les estimations qui en découlent sont en général différentes, mais qu'aucune n'est systématiquement préférable à l'autre.
L'utilisation des formules données ci-dessus nécessite la détermination de deux quantités mais lesquelles ? En pratique, le calcul exact de ces quantités est très complexe pour le type d'échantillonnage introduit dans la première partie. Que proposez-vous alors pour obtenir ces quantités ?

PARTIE III

Pour évaluer l'erreur commise en traitant un échantillon tiré à PISR par des méthodes d'analyse plus simples, les données ont été traitées comme si elles provenaient , d'une part, d'un échantillonnage à probabilités égales sans remise (PESR), et, d'autre part, d'un échantillonnage à probabilités inégales avec remise (PIAR).

1. Rappeler l'estimateur du total dans le premier cas, c'est-à-dire, l'échantillonnage à PESR, que l'on note $\widehat{Y}_2$.
2. Rappeler la formule de la variance de cet estimateur ainsi que l'estimateur de cette dernière.
3. Rappeler l'estimateur du total dans le second cas, c'est-à-dire, l'échantillonnage à PIAR, que l'on note $\widehat{Y}_3$. On introduira, si nécessaire, la probabilité de sélection que l'on notera z_i. D'ailleurs, pouvez-vous expliquer comment vous construisez cette probabilité z_i dans la pratique ?
4. Rappeler la formule de la variance de cet estimateur ainsi que l'estimateur de cette dernière.

PARTIE IV

Que conclure de cette étude ? On pourra PAR EXEMPLE envisager de calculer les précisions de chaque estimateur et de les comparer entre elles.

Cet examen s'inspire de l'article de F. Bertrand et R. Palm, Échantillonnage avec probabilités inégales et sans remise : comparaison de l'estimateur de Horvitz et Thompson à des alternatives plus simples, Revue de statistique appliquée, tome 38, numéro 3, 1990, p.5-21.

BIBLIOGRAPHIE

– GOURIEROUX C. (1981). *Théorie des sondages.* Paris, Economica, p.272
– HORVITZ D.G. et THOMPSON D.J. (1952) A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Ass.,* 47, p.663-685
– YATES F. et GRUNDY P.M. (1953) Selection without replacement from within strata with probabilities proportional to size. *J. Roy. Statist. Soc., Series B,* 15, p.253-261