
TD 6 : Utilisation d'une information auxiliaire

Exercice 1. Soit un plan stratifié composé de H strates de taille N_h . On souhaite estimer la moyenne de la population μ_y d'un caractère y . Notons $\mu_{x,h}$, $h = 1, \dots, H$, les moyennes dans les strates d'un caractère auxiliaire x qui sont supposées connues.

Votre chef propose d'estimer μ_y via l'estimateur suivant

$$\tilde{\mu}_y = \hat{\mu}_y + \mu_x - \hat{\mu}_x,$$

où $\hat{\mu}_y$ et $\hat{\mu}_x$ sont les π -estimateurs μ_y et μ_x respectivement.

On réalise un sondage aléatoire simple, sans remise dans chaque strate.

1. Montrez que $\tilde{\mu}_y$ est un estimateur sans biais de μ_y .
2. Donnez sa variance.
3. Quelle est l'allocation optimale des n_h pour minimiser sa variance? On négligera le facteur de correction en population finie.
4. Quand est-ce que $\tilde{\mu}_y$ est préférable à $\hat{\mu}_y$?



Exercice 2. Deux dentistes font une enquête sur l'état des dents des 200 enfants d'un village. Le premier dentiste sélectionne selon un sondage aléatoire simple 20 enfants parmi les 200, et comptabilise les effectifs dans l'échantillon selon le nombre de dents cariées. Les résultats sont présentés au sein du tableau ci-dessous.

Nombre de dents cariées	0	1	2	3	4	5	6	7	8
Nombre d'enfants	8	4	2	2	1	2	0	0	1

Le second dentiste examine les 200 enfants, mais dans le seul but de déterminer ceux qui n'ont aucune carie. Il constate que 50 enfants sont dans ce cas.

1. Estimez le nombre moyen de dents cariées par enfant dans le village en utilisant seulement les résultats du premier dentiste. Estimez la précision de l'estimateur obtenu et l'intervalle de confiance associé.
2. Proposez un autre estimateur du nombre moyen de dents cariées par enfant en utilisant les résultats des deux dentistes. Calculez la nouvelle estimation et appréciez le gain d'efficacité obtenu.



Exercice 3. Le directeur d'une entreprise de confection de chaussure veut estimer la longueur moyenne des pieds droits des hommes adultes d'une ville. Soient y le caractère « longueur du pied droit » (en cm) et x la taille de l'individu (en cm).

Le directeur sait en outre par les résultats d'un recensement que la taille moyenne des hommes adultes de cette ville est de 168cm. Pour estimer la longueur des pieds, le directeur effectue un sondage aléatoire simple sans remise de 100 hommes adultes. Les résultats sont les suivants :

$$\bar{x} = 169, \quad \bar{y} = 24, \quad s_{xy} = 15, \quad s_x^2 = 100, \quad s_y^2 = 4.$$

Sachant que 400000 hommes adultes vivent dans cette ville,

1. Calculez le π -estimateur, l'estimateur par le quotient, l'estimateur par différence et l'estimateur par la régression.
2. Estimez les variances de ces 4 estimateurs.
3. Quel estimateur conseilleriez-vous au directeur ?
4. Exprimez la différence entre la variance estimée de l'estimateur par le quotient et la variance estimée de l'estimateur par la régression, en fonction de \bar{x} , \bar{y} , de la pente \hat{a} de la régression de y sur x . Commentez.