

So - M - easy

So function

$$1 - \text{Var}(GA) = \left(1 - \frac{n}{N}\right) = \frac{S^2}{n}$$

$$= \left(1 - \frac{10}{20}\right) = \frac{MS_{305}}{10}$$

$$= 5,77$$

$$2 - S_{y,h} = \frac{1}{N_h - 1} \sum_{k \in U_h} (y_k - \mu_{y,h})^2$$

$$\mu_1 = 22,9$$

$$\mu_2 = 44,83$$

$$\mu_3 = 38,5$$

$$S_1^2 = 6,99$$

$$S_2^2 = 26,17$$

$$S_3^2 = 6,33$$

* On en déduit la variance de l'estimateur

$$\text{Var}(\hat{\mu}) = \frac{1}{n} \left(1 - \frac{n}{N}\right) \frac{1}{N} \sum_{h=1}^H N_h S_h^2$$

$$= \frac{1}{10} \left(1 - \frac{10}{20}\right) \frac{1}{20} [10 \times 6,99 + 6 \times 26,17 + 4 \times 6,33]$$

$$= 0,63$$

3 - Variance de l'estimateur est de 4,86.

V - PLAN STRATIFIÉ OPTIMAL DU LT TOTALE

Si votre intérêt est d'estimer un total ou une moyenne, alors il existe une taille optimale pour les strates. On cherche donc les tailles d'échantillon $n_1, \dots, n_h$ minimisant la variance de l'estimateur du total ty pour une

taille d'échantillon fixée n , c-a-d minimiser la variance.

$$\text{Var}(y_{\text{strat}}) = \sum_{h=1}^H N_h (N_h - n_h) \frac{S_{y,h}^2}{n_h}$$

par rapport à n_h et sous contrainte de

$$\sum_{h=1}^H n_h = n$$

le Lagrangien de ce problème de minimisation est :

$$\mathcal{L}(n_1, \dots, n_H, \lambda) = \sum_{h=1}^H N_h (N_h - n_h) \frac{S_{y,h}^2}{n_h} + \lambda \left(\sum_{h=1}^H n_h - n \right)$$

On a donc :

$$\frac{\partial \mathcal{L}}{\partial n_h} = 0 \Rightarrow -\frac{N_h^2}{n_h^2} S_{y,h}^2 + \lambda = 0$$

$$\Rightarrow n_h = \frac{N_h S_{y,h}}{\sqrt{\lambda}}$$

Mais puis que $\sum n_h = n$, on a donc

$$\lambda^{-\frac{1}{2}} \sum_{h=1}^H N_h S_{y,h} = n \quad \lambda^{-\frac{1}{2}} \sum_{h=1}^H N_h S_{y,h} = n$$

et il vient :

$$n_h = n \frac{N_h S_{y,h}}{\sum_{j=1}^H N_j S_{y,j}}$$

Délégué Nsangou Mohamed

$$h = 1, \dots, H$$

Remarque : la taille optimale pour la strate U_h est donc proportionnelle au produit de la taille de cette strate et de l'écart type du caractère y de cette strate.

Bien entendu en pratique, on ne connaît pas $S_{y,h}$ et donc la formule précédente n'est pas d'un grand intérêt. Elle est cependant assez instructive et intuitive ! Instructive puisque elle indique qu'il faut surreprésenter les strates qui ont une forte variabilité.

Remarque : En pratique les tailles $n_{y,h}$ et on arrondira les résultats. De plus il peut arriver également que $n_{y,h} > N_h$ pour A_h un $h \in h = 1, \dots, H$. Pour de tels cas, on posera alors $n_{y,h} = N_h$ et on déterminera les tailles optimales des strates restantes en procédant par itération si nécessaire.

Supposons que nos tailles optimales soient des entiers et telles que $n_{y,h} < N_h \forall h$. Alors la variance du $\bar{y}$ estimateur est donc :

$$\text{Var}(\hat{y}_{\text{strat}}) = \frac{1}{n} \sum_{h=1}^H N_h (N_h - n_h) \frac{s_{y,h}^2}{N_h}$$

$$= \frac{1}{n} \sum_{h=1}^H N_h^2 \frac{\sum_{i=1}^H N_h s_{y,i}^2}{n N_h s_{y,h}^2} - \frac{1}{n} \sum_{h=1}^H N_h s_{y,h}^2$$

$$= \left(\frac{\sum_{h=1}^H N_h s_{y,h}^2}{n} \right) \frac{1}{n} \sum_{h=1}^H N_h s_{y,h}^2 - \frac{1}{n} \sum_{h=1}^H N_h s_{y,h}^2$$

$$= \frac{1}{n} \left(\sum_{h=1}^H N_h s_{y,h}^2 \right) - \frac{1}{n} \sum_{h=1}^H N_h s_{y,h}^2$$