

est donné par :

$$\text{Var}(\hat{t}_{y,\pi}) = \frac{M-m}{m-1} \frac{M}{m} \sum_{i \in \mathcal{G}} \left(t_{g,i} - \frac{\hat{t}_y}{M} \right)^2$$

11/12/24

*. tirage proportionnel aux tailles des grappes : l'expression précédente montre que $\text{Var}(\hat{t}_{y,\pi})$ ie la précision de l'estimateur dépend du total par grappe ie du choix des grappes ainsi que des probabilités d'inclusion $\pi_{g,i}$; On établit ainsi que la majeure partie de la variance dans le cas des tirages à plusieurs degrés provient souvent du 1^{er} tirage, un moyen pour que la variance soit minimale est de s'assurer que le rapport $\frac{t_{y,i}}{\pi_{g,i}}$ ne dépende pas de la grappe i ; ie Condition est approximativement réalisée lorsque les

groupes sont hétérogènes pour le caractère y (peut de ressemblance entre les unités primaires) et que les probabilités d'inclusions sont proportionnelles au taille des groupes, en effet les sous ensembles regroupés dans une unité primaire ont souvent tendance à se ressembler.

Ainsi pour palier à cet effet on effectue un tirage proportionnelle à la taille des groupes; On suppose ici que le nbre de groupes de l'échantillon est donné par la relation

$$mN_i \leq N \quad i \in \{1, \dots, M\}$$

les probabilités d'inclusion de groupes sont alors

$$\pi_{g,i} = \frac{mN_i}{N}$$

Conditions, la taille n_g de l'échantillon S vaut en moyenne :

$$E(n_g) = E\left(\sum_{i \in S} N_i\right) = \sum_{i=1}^M N_i \pi_{g,i} = \frac{m}{N} \sum_{i=1}^M N_i^2$$

Les II-estimateurs du total et de la moyenne sont donnés par :

$$\hat{t}_{y,\pi} = \frac{N}{m} \sum_{i \in S_y} \frac{t_{y,i}}{N_i} = \frac{N}{m} \sum_{i \in S_y} \mu_{y,i}$$

$$\hat{\mu}_{y,\pi} = \frac{1}{m} \sum_{i \in S_y} \mu_{y,i}$$

L'estimateur du total est sans biais de variance pour un plan à taille fixe, donné par

$$\text{Var}(\hat{t}_{y,\pi}) = -\frac{1}{2} \sum_{i,j=1}^M \left(\frac{t_{y,i}}{N_i} - \frac{t_{y,j}}{N_j} \right)$$

$$= -\frac{1}{2} \sum_{\substack{i,j=1 \\ i \neq j}}^M \left(\frac{N t_{y,i}}{N_i} - \frac{N t_{y,j}}{N_j} \right) \left(\frac{t_{y,i}}{N_i} - \frac{t_{y,j}}{N_j} \right)$$

$$= -\frac{N^2}{2m^2} \sum_{\substack{i,j=1 \\ i \neq j}}^M (\mu_{y,i} - \mu_{y,j}) \left(\frac{N t_{y,i}}{N_i} - \frac{N t_{y,j}}{N_j} \right) \left(\frac{t_{y,i}}{N_i} - \frac{t_{y,j}}{N_j} \right)$$

$$\frac{m^2}{N^2} N_i \cdot N_j$$

6-1) 2^e effet de grappe

Les plans par grappes induisent très souvent une perte de précision liée au fait que des similarités existent entre les individus (US: Unité Sociale) d'une m (UP: Unité Primitif) par exple une enquête sur les revenus dans les districts d'une commune peut être affecté par la similarité des revenus en fonction des districts car les districts favorisés regroupent probablement, les ménages au haut revenus et les districts du Centre Ville présente une forte concentration de Commerce.

Une mesure de cet effet de grappe est donné par le coefficient de corrélation intra grappe noté ρ et exprimé pour un tirage à 2

dégré's par :

$$P = \frac{\sum_{i=1}^M \sum_{\substack{k \neq l \\ k, l=1 \\ k, l=N}}^{N_i} (y_k - \mu_y)(y_l - \mu_y)}{\sum_{k=1}^{N_i} (y_k - \mu_y)^2} \cdot \frac{1}{N-1}$$

ou μ est la taille moyenne des (VF)

ainsi si il y a une forte similarité entre les individus k et l ($k \neq l$) le produit $(y_k - \mu_y)(y_l - \mu_y)$ sera positif.

Si par contre ces 02 individus sont situés de part et d'autre de la moyenne général μ_y alors l'ap contient des individus très différents et le produit $(y_k - \mu_y)(y_l - \mu_y)$ sera négatif.

A partir de la valeur de P on peut sous certaines conditions calculer la précision d'un son

lage a 2 degrés. En effet on
 suppose si l'on effectue l'estimateur
 d'un total par un tirage aléatoire
 simple a 02 degrés lorsque toute
 les **UP** sont de m taille et l'
 échantillon d'**US** est de taille
 constante $\bar{N}$ alors on a:

$$\text{Var}(\hat{t}_y) = \sigma_y^2 = \sigma_{SAS}^2 (1 + f(\bar{N} - 1))$$

* Cas particulier

Dans le cas d'un sondage
 en grappe $\bar{N} = \bar{N}_i$

$$\text{Var}(\hat{t}_y) = \text{Var}(\hat{t}_{SAS}) \cdot (1 + f(\bar{N}_i - 1))$$

où $\sigma_{SAS}^2 = \text{Var}(\hat{t}_{SAS})$ est la vari-
 ance de l'estimateur de m total
 a partir d'un plan de sondage
 aléatoire simple **CO. abrégé SAS**
 la grandeur **ΔEFF** obtenue
 par le rapport des 02 Variances

d'estimation permet d'estimer la perte de ~~précision~~ **précision obtenue** lors du passage d'un plan de sondage à l'autre. On appelle **effet de sondage (Design effect)** il est noté: ~~AE~~

$$AE = 1 + p(N-1) = \frac{\text{Var}(\hat{y})}{\text{Var}(\hat{y}_{SAS})}$$

Ce qui implique que

$$\text{Var}(\hat{y}) < \text{Var}(\hat{y}_{SAS}) \text{ si } p < 1$$

ie qu'il y a une grande hétérogénéité (Variation) à l'intérieur de la grappe. **Exple:** Considérons une population de **6 individus** donc les valeurs des caractères y sont les suivantes:

Ind	y	$(y_i - \bar{y})^2$
1	$y_1 = 1$	81
2	$y_2 = 2$	64
3	$y_3 = 4$	36
4	$y_4 = 6$	4
5	$y_5 = 15$	25
6	$y_6 = 30$	400
total	$\sum y = 60$	640

1) Calculer la moyenne générale μ_y et la variance corrigée S_y^2

Solut°

$$\mu_y = \frac{\sum y_i}{N} = \frac{60}{10} = 10$$

(Variance corrigée) $\leftarrow S_y^2 = \frac{1}{N-1} \sum (y_i - \mu_y)^2$

$$S_y^2 = \frac{640}{5} = 128$$

2) on constitue deux groupes selon:
2 configurations

1^{re} configuration $DP_1 = \{1, 2, 3\}$, $DP_2 = \{4, 5\}$

2^{em} configuration $DP_1 = \{1, 4, 6\}$, $DP_2 = \{2, 3\}$

Calculer le coefficient de corrélation intra groupe et inter groupe:

Solut° on va deabord extraire les couples dans chaque cas ; 202 identique.

pour la 1^{re} configuration $VP_1 \Rightarrow \{(1,2), (1,3), (2,3)\}$

$VP_2 \Rightarrow \{(4,5), (4,6), (5,6)\} \times 2$

pour la 2^e - 11 - on a : $\{(1,4), (1,6), (4,6), (2,3), (2,5), (3,5)\}$

On a alors :

$$r = \frac{1}{N-1} \cdot \frac{\sum_{k=1}^M \sum_{j=1}^M (y_k - \mu_y)(y_j - \mu_y)}{\sum_{k=1}^N (y_k - \mu_y)^2} \quad \text{avec } N=3$$

1^{re} configuration :

$$= \frac{1}{3-1} \cdot \left(2[(1-10)(2-10) + (1-10)(4-10) + (2-10)(4-10)] + 2[(8-10)(15-10) + (8-10)(30-10) + (15-10)(30-10)] \right)$$

690

$\rho = 0,366$ dans la 1^{re} Configurat°

le m calcul dans la 2^{de} Configurat° permet de déterminer $\rho = -0,366$

Conclusion Conformément à l'intuit° la 2^{de} Configurat° est préférable à la 1^{re} le signe moins de ρ traduisant une diminution de l'effet de grappe ou l'hétérogénéité des individus au sein d'une grappe

II-) Le sondage à 2 degrés

Le sondage à 2 degrés qui se son nom l'indique sur les VP d'abord et les VP en suite de la m manière que précédemment la populat° est divisée en 2 sous population V_i qui sont elle m composés de $N_i = 10 = 15$

de sorte qu'on retrouve $\sum_{i=1}^M M_i = N$

Ainsi pour effectuer un plan à 2 degrés il faut construire un échantillon S_1 d'UP à partir d'un plan de sondage $P_1(\cdot)$ puis pour chaque UP sélectionnée construire un échantillon S_2 sur les US à partir d'un plan de sondage $P_2(\cdot)$ d'une manière générale, est souhaitable que le plan du 2^{ème} degré soit très indépendant du 1^{er} plan $P_1(\cdot)$ tel que

$$P_2(S_2 = s_2 / S_1 = s_1) = P_1(S_2 = s_2)$$

les tirages du 2^{ème} degré doivent également être mutuellement indépendants

1) probabilités d'inclusion

Notons $\pi_{1,i}$ et $\pi_{1,ij}$ les probabilités d'inclusion d'ordre 1 et 2 pour le 1er degré ie

$$\pi_{1,i} = P_n(V_i \in S_n)$$

$$\pi_{1,ij} = P_n(V_i \in S_n, V_j \in S_n)$$

Notons également $\pi_{R/i}$ la probabilité que U_{S_R} soit sélection sachant que U_{P_i} a été choisi et $\pi_{Rl/i}$ la probabilité d'inclusion d'ordre 2 de $R(l)$ sachant que V_i a été retenue. On a alors

$$\pi_{R/i} = P(R \in S_n | i \in S_n) P(i \in S_n) = \pi_{R/i} \pi_{1,i}$$

$$\text{et } \pi_{Rl/i} = \begin{cases} \pi_{Rl/i} + \pi_{1,i} & R, l \in V_i \\ \pi_{Rl/i} + \pi_{Rl/j} + \pi_{1,ij} & R, l, i \in V_j, R \neq l \end{cases}$$

$$\Delta_{Rl/i} = \pi_{Rl/i} - (\pi_{R/i} - \pi_{Rl/j})$$

2.) le π -estimateur

Ce précédemment le total t_y
écrit : $t_y = \sum_{i=1}^n t_{y,i}$ avec $t_{y,i} = \sum_{k \in S_i} y_k$

le π -estimateur de ce total est
alors

$$\begin{aligned} \hat{t}_{y,\pi} &= \sum_{i \in S_n} \sum_{k \in S_i} \frac{y_k}{\pi_{k,i} - \pi_{1,i}} \\ &= \sum_{i \in S_n} \frac{t_{y,i}}{\pi_{1,i}} \quad (\text{avec } \hat{t}_{y,i} = \sum_{k \in S_i} \frac{y_k}{\pi_{k,i}}) \end{aligned}$$

Cet estimateur est sans biais
de variance égale à :

$$\text{Var}(\hat{t}_{y,\pi}) = M^2 \left(1 - \frac{m}{M}\right) \frac{S_1^2}{m} + \frac{M}{m} \sum_{i=1}^n$$

$$N_i^2 \left(1 - \frac{n_i}{N_i}\right) \frac{S_{2,i}}{n_i}$$

$$= \underbrace{\text{Var}_{UP}}_{\text{inter centre UP}} + \underbrace{\text{Var}_{US}}_{\text{intra centre US}}$$

inter
centre
UP)

intra
centre
US)

avec $S^2 = \frac{1}{M-1} \sum_{i=1}^M (y_i - \bar{y})^2$

et $S_{2,i}^2 = \frac{1}{N_i-1} \sum_{R \in U_i} (y_{i,R} - \bar{y}_i)^2$

un estimateur de cette variance est donné par

$$\hat{V}(\hat{y}, \pi) = M^2 \left(1 - \frac{m}{M}\right) \cdot \frac{\sum_{i=1}^M S_i^2}{m} + \frac{M}{m} \sum_{i \in S} \left(\frac{N_i^2}{n_i} \left(1 - \frac{n_i}{N_i}\right) S_{2,i}^2 \right)$$

Rmq:

→ Si tout les UP sont tirés avec une probabilité certaine (égale 1) Alors $VAR(UP) = 0$ et on retrouve le sondage stratifié donc les strates correspond au UP

- Si toute les US sont tirés avec probabilités Certainne alors ~~Non~~ (US) et on retrouve le Sondage par grappe.

Exercice

L'objectif d'un sondeur est d'estimer le Revenu moyen des ménages dans une ville Compose de 60 quartiers par cela il selectonne 03 quartiers par un sondage aléatoire simple et interroge tous les ménages qui y reside, il sait par ailleurs que 5000 ménage reside dans cette ville, le Resumé de son enquête est donné dans le tableau suit :

N du quartier	N de ménage	Revenu
1	120	
2	100	
3	80	

1) Quel est le revenu moyen $\bar{y}$ et le revenu total $\sum y_i$ selon la méthode de dans la ville considérée

2.) Quel est la Variance estimer sans biais de l'estimateur de Robit Stinson de la moyenne?

Solut°