



Université PANTHÉON - ASSAS (PARIS II)

U.E.C.1

Droit - Economie - Sciences Sociales

4615

Session : Janvier 2023

Année d'étude : Première année de Master économie-gestion mention ingénierie économique et statistique

Discipline : **Sondages** (Unité d'Enseignements Complémentaires 1)

Titulaire(s) du cours : M. Philippe PERIE

Document(s) autorisé(s) : Formulaire

Les exercices sont indépendants et peuvent être résolus dans n'importe quel ordre.
Barème indicatif : questions : 8 points, exercice 1 : 4 points, exercice 2 : 10 points

Questions :

1. Que pensez-vous des affirmations suivantes
 - a. Un sondage stratifié, quelle que soit son critère de stratification donnera toujours des estimations au moins aussi précises qu'un sondage aléatoire de même taille : vrai / faux
 - b. Un sondage à probabilité inégale, quelle que soit son critère de définition des probabilités de sélection des individus, donnera toujours des estimations au moins aussi précises qu'un sondage aléatoire de même taille : vrai / faux
2. L'échantillonnage non probabiliste est un moyen de sélectionner des unités d'une population à l'aide d'une méthode subjective (c.-à-d. non aléatoire). Il n'est pas nécessaire d'avoir une base de sondage complète pour l'échantillonnage non probabiliste. Quels sont les avantages et les limites de cette méthode par rapport à un sondage probabiliste ?
3. Donner un plan de sondage qui permet de sélectionner à probabilités inégales, mais avec une meilleure précision qu'un sondage aléatoire simple
4. Qu'est-ce qui peut justifier selon vous l'utilisation de plans en plusieurs degrés ? Donner un exemple

Exercice 1 : On souhaite estimer la quantité d'eau moyenne (exprimée en m³) consommée annuellement par les habitants d'une ville donnée de 100 000 habitants. On sélectionne par un plan simple un échantillon de 250 habitants.

Les résultats obtenus sont les suivants :

$$\sum_{i=1}^n x_i = 15\,125$$

$$\sum_{i=1}^n x_i^2 = 921\,310$$

1. Donner un intervalle de confiance à 95% pour la quantité d'eau moyenne consommée annuellement par les habitants de cette ville.
2. On s'intéresse maintenant à la quantité totale consommée annuellement par l'ensemble des habitants de la ville. Donner une estimation, puis un intervalle de confiance à 95% pour cette quantité totale.

Exercice 2 : Un collectif européen de cirques abandonne les spectacles d'animaux sauvages et se retrouve avec 100 éléphants à rapatrier en Afrique par avion depuis une ville d'Europe. Les éléphants sont classés en deux catégories : les mâles et les femelles. Afin d'affréter les équipages et les avions, on veut estimer le poids total du troupeau. L'année précédente, on avait fait peser les 100 éléphants et on avait trouvé ceci.

	Effectifs N_h	Moyennes $\widehat{\mu}_h$	$\widehat{\sigma}_h^2$
Mâles	60	6	4
Femelles	40	4	2.25

Le problème va être d'estimer le plus précisément possible le poids du troupeau de 100 car affréter un avion-cargo coûte cher, et peser 100 éléphants est compliqué et on a ni le matériel ni le temps ... Chaque avion peut porter 50 tonnes de cargo, on veut limiter le nombre de voyages.

1. Calculer σ^2 et $\widehat{\sigma}^2$ pour l'année précédente. (utiliser la formule de la décomposition de la variance totale)

$$\sigma^2 = \frac{1}{N} \sum_{h=1}^H N_h \sigma_h^2 + \frac{1}{N} \sum_{h=1}^H N_h (\mu_h - \mu)^2 = \sigma_{\text{intra}}^2 + \sigma_{\text{inter}}^2.$$

2. Le directeur suppose désormais que les dispersions de poids n'évoluent pas sensiblement d'une année sur l'autre (ce type d'hypothèse reste ici très raisonnable et se rencontre couramment en pratique quand on répète des enquêtes dans le temps). Si le directeur procède à un tirage aléatoire simple de 10 éléphants, quelle est la variance de l'estimateur du poids total du troupeau ?
3. Si le directeur procède à un tirage stratifié avec allocation proportionnelle de 10 éléphants, quelles tailles d'échantillon doit-on retenir dans chaque strate ? Quelle est alors la variance de l'estimateur du poids total du troupeau ?
4. Si le directeur procède à un tirage stratifié optimal de 10 éléphants, quelles tailles d'échantillon doit-on retenir dans chaque strate ? Quelle est alors la variance de l'estimateur du poids total du troupeau ?
5. Calculer les estimations ponctuelles et les intervalles de confiance à 95% dans les deux cas des sondages stratifiés. Combien d'avions à affréter dans chaque cas pour être sûrs à 95% de pouvoir rapatrier tous les éléphants ? Commenter

Master 1 - Economie-Gestion Mention Ingénierie économique et statistique 2022 - 2023

Formulaire - Techniques de Sondage



Ce document est un formulaire accompagnant le cours 'Techniques de Sondage' du Master 1 IES, il est autorisé pour l'examen.

Philippe Périé

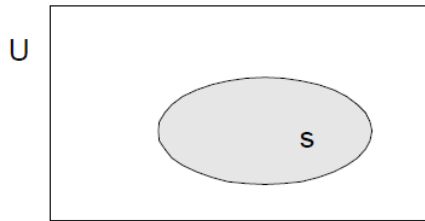
Contents

Plans aléatoires simples	2
Plans a probabilités inégales	4
Plans stratifiés	5
Plans par grappes	7
Plans a deux degrés	9
Redressements	11

1. Plans aléatoires simples

Principe

Tirage d'un échantillon de n unités sans remise et à probabilités égales dans une population de taille N finie



Notations

1. Dans la population (ou univers) $U = \{1, 2, \dots, k, \dots, N\}$

- Variable d'intérêt : Y de caractéristique individuelle Y_k
- Total : $T_Y = \sum_{k \in U} Y_k$
- Moyenne : $\bar{Y} = \frac{T_Y}{N} = \frac{1}{N} \sum_{k \in U} Y_k$
- Variance : $\sigma_y^2 = \frac{1}{N} \sum_{k \in U} (Y_k - \bar{Y})^2$
- Dispersion (variance modifiée) : $S_y^2 = \frac{1}{N-1} \sum_{k \in U} (Y_k - \bar{Y})^2 = \frac{N}{N-1} \sigma_y^2$

2. Dans l'échantillon s : sous-ensemble de U de taille $n(s)$

- Ensemble des échantillons possibles : S
 - Plan de sondage probabiliste : loi de probabilité sur S
- $$p(s) \geq 0, \forall s \in S, \text{ et } \sum_{s \in S} p(s) = 1.$$
- Moyenne : $\hat{y} = \frac{1}{n} \sum_{k \in s} Y_k$
 - Dispersion empirique : $\hat{s}_y^2 = \frac{1}{n-1} \sum_{k \in s} (Y_k - \hat{y})^2$
 - Probabilité d'inclusion d'ordre un de k : $\pi_k = P(k \in s) = \sum_{s \in S / k \in s} p(s)$
 - Probabilité d'inclusion ou double de k et l : $\pi_{kl} = P(k \in s, l \in s) = \sum_{s \in S / k, l \in s} p(s)$
 - $\Delta_{kl} = \pi_{kl} - \pi_k \pi_l$

Formulaire sondage aléatoire simple

Probabilité de sélectionner l'échantillon s : $p(s) = 1/C_N^n$

Probabilité de sélectionner l'individu k : $\forall k \in U, \pi_k = P(k \in s) = \frac{n}{N} = f$ (taux de sondage)

Paramètre d'intérêt Statistique	Moyenne	Proportion $p = N_0/N$	Total
Estimateur du paramètre d'intérêt	$\hat{y} = \frac{1}{n} \sum_{k \in S} Y_k = \hat{y}(s)$	$\hat{p} = \frac{1}{n} \sum_{k \in s} y_k = \frac{n_0}{n}$	$\hat{t}_y = N \times \hat{y} = \frac{N}{n} \sum_{k \in S} Y_k$
Vraie variance d'échantillonnage de cet estimateur	$Var(\hat{y}) = \left(1 - \frac{n}{N}\right) \frac{S_y^2}{n}$	$Var(\hat{p}) = \left(1 - \frac{n}{N}\right) \frac{N}{N-1} \frac{p(1-p)}{n}$	$Var(\hat{t}_y) = N^2 \left(1 - \frac{n}{N}\right) \frac{S_y^2}{n}$
Estimateur de la variance d'échantillonnage	$\hat{Var}(\hat{y}) = \left(1 - \frac{n}{N}\right) \frac{\hat{S}_y^2}{n}$	$\hat{Var}(\hat{p}) = \left(1 - \frac{n}{N}\right) \frac{\hat{p}(1-\hat{p})}{n-1}$	$\hat{Var}(\hat{t}_y) = N^2 \left(1 - \frac{n}{N}\right) \frac{\hat{S}_y^2}{n}$

Intervalle au niveau de confiance 95% pour la moyenne :

$$IC_{95\%}(\bar{Y}) = \left[\hat{y} - 1,96\sqrt{\hat{Var}(\hat{y})}, \hat{y} + 1,96\sqrt{\hat{Var}(\hat{y})} \right]$$

sous hypothèse que n est grand $\frac{\hat{y} - \bar{Y}}{\sqrt{\hat{Var}(\hat{y})}} \rightarrow N(0, 1)$

2. Plans a probabilités inégales

Principe

Retenir les unités les plus porteuses d'information : sélection des unités avec des probabilités proportionnelles à une variable X de taille corrélée positivement à une variable d'intérêt Y

Notations

Formulaire

Probabilité de sélectionner l'individu k :

- Pour un plan à probabilités proportionnelles à une variable X de taille (corrélée positivement à Y)

$$\forall k \in U, \pi_k = P(k \in S) = n \frac{X_k}{\sum_{k \in U} X_k}$$

- Pour un plan de taille fixe, $\sum_{k \in U} \pi_k = n$

Statistique \ Paramètre d'intérêt	Moyenne	Total
Estimateur d'Horvitz-Thompson du paramètre d'intérêt (π-estimateur)	Si la taille N est connue : $\hat{\mu}_{y\pi} = \frac{1}{N} \sum_{k \in S} \frac{Y_k}{\pi_k} = \frac{\hat{t}_{y\pi}}{N}$ Sinon, estimateur de Hájek : $\hat{\mu}_{yH} = \frac{1}{\hat{N}_\pi} \sum_{k \in S} \frac{Y_k}{\pi_k} = \frac{\sum_{k \in S} \frac{Y_k}{\pi_k}}{\sum_{k \in S} \frac{1}{\pi_k}} = \frac{\hat{t}_{y\pi}}{\hat{N}_\pi}$	$\hat{t}_{y\pi} = \sum_{k \in S} \frac{Y_k}{\pi_k}$ En particulier : $\hat{N}_\pi = \sum_{k \in S} \frac{1}{\pi_k}$
Vraie variance d'échantillonnage de cet estimateur	Cas général $Var(\hat{\mu}_{y\pi}) = \frac{1}{N^2} \sum_{k \in U} \sum_{l \in U} \frac{Y_k}{\pi_k} \frac{Y_l}{\pi_l} \Delta_{kl}$ Si la taille de l'échantillon est fixe $Var(\hat{\mu}_{y\pi}) = \frac{1}{2N^2} \sum_{k \in U} \sum_{l \in U} \left(\frac{Y_k}{\pi_k} - \frac{Y_l}{\pi_l} \right)^2 \Delta_{kl}$	Cas général : $Var(\hat{t}_{y\pi}) = \sum_{k \in U} \sum_{l \in U} \frac{Y_k}{\pi_k} \frac{Y_l}{\pi_l} \Delta_{kl}$ Si la taille de l'échantillon est fixe $Var(\hat{t}_{y\pi}) = \frac{1}{2} \sum_{k \in U} \sum_{l \in U} \left(\frac{Y_k}{\pi_k} - \frac{Y_l}{\pi_l} \right)^2 \Delta_{kl}$
Estimateur de la variance d'échantillonnage	Cas général $\hat{Var}(\hat{\mu}_{y\pi}) = \frac{1}{N^2} \sum_{k \in S} \sum_{l \in S} \frac{Y_k}{\pi_k} \frac{Y_l}{\pi_l} \frac{\Delta_{kl}}{\pi_{kl}}$ Si la taille de l'échantillon est fixe $\hat{Var}_2(\hat{\mu}_{y\pi}) = \frac{1}{2N^2} \sum_{k \in S} \sum_{l \in S} \left(\frac{Y_k}{\pi_k} - \frac{Y_l}{\pi_l} \right)^2 \frac{\Delta_{kl}}{\pi_{kl}}$	Cas général $\hat{Var}(\hat{t}_{y\pi}) = \sum_{k \in S} \sum_{l \in S} \frac{Y_k}{\pi_k} \frac{Y_l}{\pi_l} \frac{\Delta_{kl}}{\pi_{kl}}$ Si la taille de l'échantillon est fixe $\hat{Var}_2(\hat{t}_{y\pi}) = \frac{1}{2} \sum_{k \in S} \sum_{l \in S} \left(\frac{Y_k}{\pi_k} - \frac{Y_l}{\pi_l} \right)^2 \frac{\Delta_{kl}}{\pi_{kl}}$

Si n est grand, l'intervalle de confiance pour la moyenne au niveau de confiance $1 - \alpha$ est :

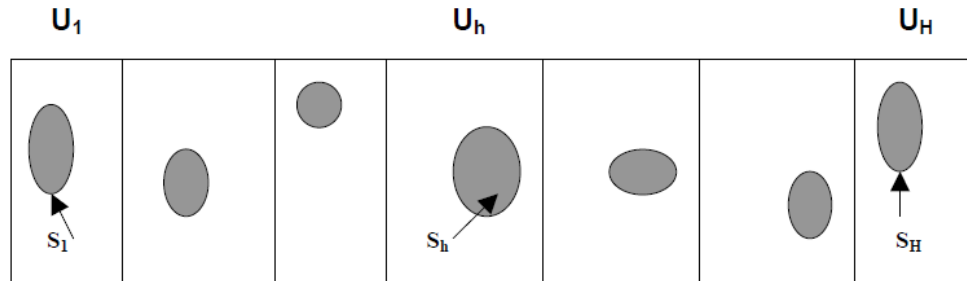
$$IC_{1-\alpha}(\mu_y) = \left[\hat{\mu}_{y\pi} - u_{1-\frac{\alpha}{2}} \sqrt{\hat{Var}(\hat{\mu}_{y\pi})}, \hat{\mu}_{y\pi} + u_{1-\frac{\alpha}{2}} \sqrt{\hat{Var}(\hat{\mu}_{y\pi})} \right]$$

où $u_{1-\frac{\alpha}{2}}$ désigne le fractile d'ordre $1 - \alpha/2$ de la loi N(0,1)

3. Plans stratifiés

Principe

Partitionner la population en sous-groupes homogènes appelés strates, et tirage d'autant d'échantillons indépendants qu'il y a de strates. H strates notées 1, ...h, ...H



Notations

1. Dans la population

- $U = \bigcup_{h=1}^H U_h$ et $N = \sum_{h=1}^H N_h$
- Total : $t_y = \sum_{h=1}^H t_{yh} = \sum_{h=1}^H N_h \bar{y}_h$
- Moyenne : $\bar{y} = \frac{t_y}{N} = \sum_{h=1}^H \frac{N_h}{N} \bar{y}_h$ avec $\bar{y}_h = \frac{1}{N_h} \sum_{k \in U_h} y_k$
- Variance : $\sigma_{\bar{y}}^2 = \frac{1}{N} \sum_{k \in U} (y_k - \bar{y})^2 = \sum_{h=1}^H \frac{N_h}{N} \sigma_{y_h}^2 + \sum_{h=1}^H \frac{N_h}{N} (\bar{y}_h - \bar{y})^2 = \sigma_{y_{intra}}^2 + \sigma_{y_{inter}}^2 = \frac{N-1}{N} S_{\bar{y}}^2$
avec $\sigma_{y_h}^2 = \frac{1}{N_h} \sum_{k \in U_h} (y_k - \bar{y}_h)^2$

2. Dans l'échantillon

- $S = \bigcup_{h=1}^H S_h$ et $n = \sum_{h=1}^H n_h$
- Moyenne dans S_h : $\hat{\bar{y}}_h = \frac{1}{n_h} \sum_{k \in S_h} y_k$
- Dispersion dans S_h : $\hat{S}_{y_h}^2 = \frac{1}{n_h - 1} \sum_{k \in S_h} (y_k - \hat{\bar{y}}_h)^2$

Formulaire

Paramètre d'intérêt Statistique	Moyenne	Proportion	Total
Estimateur du paramètre d'intérêt	$\hat{\bar{y}} = \sum_{h=1}^H \frac{N_h}{N} \hat{y}_h$	$\hat{p} = \sum_{h=1}^H \frac{N_h}{N} \hat{p}_h$	$\hat{t}_y = N \hat{\bar{y}} = \sum_{h=1}^H \hat{t}_{yh} = \sum_{h=1}^H N_h \hat{y}_h$
Vraie variance d'échantillonnage de cet estimateur	$Var[\hat{\bar{y}}] = Var\left[\sum_{h=1}^H \frac{N_h}{N} \hat{y}_h\right] = \sum_{h=1}^H Var\left[\frac{N_h}{N} \hat{y}_h\right]$ Si plan simple dans chaque strate : $Var[\hat{\bar{y}}] = \sum_{h=1}^H \left(\frac{N_h}{N}\right)^2 \left(1 - \frac{n_h}{N_h}\right) \frac{S_{yh}^2}{n_h}$	Si plan simple dans chaque strate $Var[\hat{p}] = \sum_{h=1}^H \left(\frac{N_h}{N}\right)^2 \left(1 - \frac{n_h}{N_h}\right) \frac{N_h}{N_h - 1} \frac{p_h(1-p_h)}{n_h}$	$Var[\hat{t}_y] = Var[N \hat{\bar{y}}] = N^2 Var[\hat{\bar{y}}]$ Si plan simple dans chaque strate : $Var[\hat{t}_y] = \sum_{h=1}^H N_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{S_{yh}^2}{n_h}$
Estimateur de la variance d'échantillonnage	Si plan simple dans chaque strate $\hat{Var}[\hat{\bar{y}}] = \sum_{h=1}^H \left(\frac{N_h}{N}\right)^2 \left(1 - \frac{n_h}{N_h}\right) \frac{\hat{S}_{yh}^2}{n_h}$	Si plan simple dans chaque strate $\hat{Var}[\hat{p}] = \sum_{h=1}^H \left(\frac{N_h}{N}\right)^2 \left(1 - \frac{n_h}{N_h}\right) \frac{\hat{p}_h(1-\hat{p}_h)}{n_h - 1}$	Si plan simple dans chaque strate $\hat{Var}[\hat{t}_y] = \sum_{h=1}^H N_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{\hat{S}_{yh}^2}{n_h}$

Intervalle au niveau de confiance 95% pour la moyenne :

$$IC_{95\%}(\bar{Y}) = \left[\hat{\bar{y}} - 1,96 \sqrt{\hat{Var}(\hat{\bar{y}})}, \hat{\bar{y}} + 1,96 \sqrt{\hat{Var}(\hat{\bar{y}})} \right]$$

sous hypothèse que n est grand $\frac{\hat{\bar{y}} - \bar{Y}}{\sqrt{Var(\hat{\bar{y}})}} \rightarrow N(0, 1)$

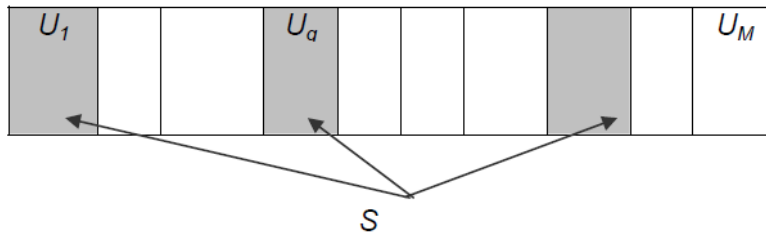
Choix des allocations

- Allocations proportionnelles : $\frac{n_h}{n} = \frac{N_h}{N} \quad \forall h \in \{1, \dots, H\}$
- Allocations optimales de Neyman (sans contrainte de budget) : $n_h = n \frac{N_h S_{yh}}{\sum_{i=1}^H N_i S_{yi}}$
- Allocations optimales sous contrainte budgétaires : $n_h = C \frac{N_h S_{yh}}{\sqrt{C_h} \sum_{i=1}^H N_i S_{yi} \sqrt{C_i}}$

4. Plans par grappes

Principe

Partition de la population en sous-groupes appelés grappes, échantillonner un certain nombre de grappes, puis recenser tous les individus d'une grappe sélectionnée



Notations

1. Dans la population U constituée de M grappes et N individus

- $U = \bigcup_{g=1}^M U_g$ et $N = \sum_{g=1}^M N_g$
- $t_y = \sum_{g=1}^M t_{yg} = \sum_{g=1}^M N_g \bar{y}_g$
- $\bar{y} = \frac{t_y}{N} = \sum_{g=1}^M \frac{N_g}{N} \bar{y}_g$ avec $\bar{y}_g = \frac{1}{N_g} \sum_{k \in U_g} y_k$
- $S_G^2 = \frac{1}{M-1} \sum_{g=1}^M \left(t_{yg} - \frac{t_y}{M} \right)^2$

2. Dans l'échantillon S constitué de m grappes et n_s individus

- $S = \bigcup_{g \in S_G} U_g$ et $n_s = \sum_{g \in S_G} N_g$

Formulaire

Paramètre d'intérêt Statistique	Total	Moyenne
Estimateur du paramètre d'intérêt	$\hat{t}_y = \frac{M}{m} \sum_{g \in S_G} t_{yg}$	$\hat{\bar{y}} = \frac{1}{N} \hat{t}_y = \frac{M}{Nm} \sum_{g \in S_G} N_g \bar{y}_g$
Vraie variance d'échantillonnage de cet estimateur	$Var[\hat{t}_y] = M^2 \left(1 - \frac{m}{M}\right) \frac{1}{m} \frac{1}{M-1} \sum_{g=1}^M \left(t_{yg} - \frac{t_y}{M}\right)^2$	$Var[\hat{\bar{y}}] = \frac{1}{N^2} Var[\hat{t}_y]$
Estimateur de la variance d'échantillonnage	$\hat{Var}[\hat{t}_y] = M^2 \left(1 - \frac{m}{M}\right) \frac{1}{m} \frac{1}{m-1} \sum_{g \in S_G} \left(t_{yg} - \frac{\hat{t}_y}{M}\right)^2$	$\hat{Var}[\hat{\bar{y}}] = \frac{1}{N^2} \hat{Var}[\hat{t}_y]$

Intervalle au niveau de confiance 95% pour la moyenne :

$$IC_{95\%}(\bar{Y}) = \left[\hat{\bar{y}} - 1,96 \sqrt{\hat{Var}(\hat{\bar{y}})}, \hat{\bar{y}} + 1,96 \sqrt{\hat{Var}(\hat{\bar{y}})} \right]$$

sous hypothèse que la taille de l'échantillon est assez grande.

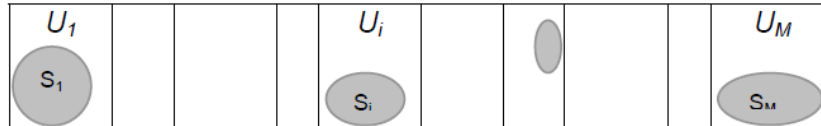
5. Plans a deux degrés

Principe

Dans une population partitionnées en sous-groupes appelés unités primaires, eux-mêmes constitués d'unités secondaires :

1^{er} degré : tirer un échantillon d'unités primaires

2^{ème} degré : tirage d'unités secondaires dans chaque unité primaire retenue, de manière indépendante



Notations

1. Dans la population U constituée de M unités primaires et N individus

- $U = \bigcup_{i=1}^M U_i$ et $N = \sum_{i=1}^M N_i$
- $t_y = \sum_{i=1}^M t_{yi} = \sum_{i=1}^M N_i \bar{y}_i$
- $\bar{y} = \frac{t_y}{N} = \sum_{i=1}^M \frac{N_i}{N} \bar{y}_i$ avec $\bar{y}_i = \frac{1}{N_i} \sum_{k \in U_i} y_k$
- $S_I^2 = \frac{1}{M-1} \sum_{i=1}^M \left(t_{yi} - \frac{t_y}{M} \right)^2$ et $S_i^2 = \frac{1}{N_i-1} \sum_{k=1}^{N_i} (y_k - \bar{y}_i)^2$

2. Dans l'échantillon S constitué de m unités primaires et n_s individus

- $S = \bigcup_{i \in SUP} S_i$ et $n_s = \sum_{i \in S_i} n_i$
- $\hat{s}_I^2 = \frac{1}{m-1} \sum_{i \in SUP} \left(\hat{t}_{yi} - \frac{\hat{t}_y}{M} \right)^2$ et $\hat{s}_i^2 = \frac{1}{n_i-1} \sum_{k \in S_i} (y_k - \hat{y}_i)^2$

Formulaire

Paramètre d'intérêt Statistique	Total	Moyenne
Estimateur du paramètre d'intérêt	$\hat{t}_y = \sum_{k \in S} \frac{y_k}{\pi_k} = \frac{M}{m} \sum_{i \in SUP} \hat{t}_{yg} = \frac{M}{m} \sum_{i \in SUP} \frac{N_i}{n_i} \sum_{k \in Si} y_k$	$\hat{\bar{y}} = \frac{1}{N} \hat{t}_y = \frac{M}{Nm} \sum_{g \in SG} N_g \hat{y}_g$
Vraie variance d'échantillonnage de cet estimateur	$Var[\hat{t}_y] = M^2 \left(1 - \frac{m}{M}\right) \frac{S_f^2}{m} + \frac{M}{m} \sum_{i=1}^M N_i^2 \left(1 - \frac{n_i}{N_i}\right) \frac{S_i^2}{n_i}$	$Var[\hat{\bar{y}}] = \frac{1}{N^2} Var[\hat{t}_y]$
Estimateur de la variance d'échantillonnage	$\hat{Var}[\hat{t}_y] = M^2 \left(1 - \frac{m}{M}\right) \frac{\hat{S}_f^2}{m} + \frac{M}{m} \sum_{i \in SUP} N_i^2 \left(1 - \frac{n_i}{N_i}\right) \frac{\hat{S}_i^2}{n_i}$	$\hat{Var}[\hat{\bar{y}}] = \frac{1}{N^2} \hat{Var}[\hat{t}_y]$

Intervalle au niveau de confiance 95% pour la moyenne :

$$IC_{95\%}(\bar{Y}) = \left[\hat{\bar{y}} - 1,96 \sqrt{\hat{Var}(\hat{\bar{y}})}, \hat{\bar{y}} + 1,96 \sqrt{\hat{Var}(\hat{\bar{y}})} \right]$$

sous hypothèse que la taille de l'échantillon est assez grande

6. Redressements

Principe

Après la collecte, intégrer de l'information auxiliaire dans la construction de l'estimateur

Formulaire (estimateur Post Stratifié et par la régression vus en cours)

En notant X l'information auxiliaire,

Méthode	Estimateur du total	Vraie erreur quadratique moyenne de cet estimateur
Estimateur d'Horvitz-Thompson	$\hat{t}_{y\pi} = \sum_{k \in S} \frac{y_k}{\pi_k} = \frac{N}{n} \sum_{k \in S} y_k$	$Var(\hat{t}_{y\pi}) = N^2 \left(1 - \frac{n}{N}\right) \frac{S_y^2}{n}$
Estimateur par la différence	$\hat{t}_{yD} = \hat{t}_{y\pi} + t_x - \hat{t}_{x\pi}$	$Var(\hat{t}_{yD}) = N^2 \left(1 - \frac{n}{N}\right) \frac{(S_y^2 + S_x^2 - 2S_{xy})}{n}$
Estimateur par le ratio (ou le quotient)	$\hat{t}_{yQ} = \hat{t}_{y\pi} \frac{t_x}{\hat{t}_{x\pi}}$	$Var(\hat{t}_{yQ}) = N^2 \left(1 - \frac{n}{N}\right) \frac{(S_y^2 + R^2 S_x^2 - 2RS_{xy})}{n}$ avec $R = \frac{t_y}{t_x} = \frac{\bar{Y}}{\bar{X}}$
Estimateur par la régression	$\hat{t}_{yD} = \hat{t}_{y\pi} + \hat{b}(t_x - \hat{t}_{x\pi})$ avec $\hat{b} = \frac{\hat{S}_{xy}}{\hat{S}_x^2}$	$Var(\hat{t}_{yQ}) = N^2 \left(1 - \frac{n}{N}\right) \frac{S_y^2}{n} (1 - \rho^2)$ avec $\rho = \frac{S_{xy}}{S_x S_y}$
Estimateur post-stratifié	$\hat{t}_{y_{post}} = \sum_{h=1}^H N_h \hat{y}_h = \sum_{h=1}^H \frac{N_h}{n_h} \sum_{k \in S_h} y_k$	$Var(\hat{t}_{y_{post}}) = \frac{N-n}{n} \sum_{h=1}^H N_h S_{y_h}^2 + \frac{N-n}{N-1} \frac{N^2}{n^2} \sum_{h=1}^H \frac{N-N_h}{N} S_{y_h}^2$

Estimateur par substitution de l'erreur quadratique moyenne

Intervalle au niveau de confiance 95% pour la moyenne :

$$IC_{95\%}(\bar{Y}) = \left[\hat{y} - 1,96\sqrt{\hat{Var}(\hat{y})}, \hat{y} + 1,96\sqrt{\hat{Var}(\hat{y})} \right]$$

sous hypothèse que la taille de l'échantillon est assez grande.