SONDAGE A DEUX DEGRÉS

- Population découpée en sous-populations dites unités primaires
- Définition: tirage de *m* unités primaires puis de n_i unités secondaires
- Avantages:
  - frais de déplacement réduits
  - absence de liste autorisée
- Mais:
  - précision moindre: effet de grappe.
  - Taille d'échantillon en général aléatoire

SONDAGE A DEUX DEGRÉS

- M unités primaires de taille N_i

$$N = \sum_{i=1}^{M} N_i \qquad T_i = \sum_{j=1}^{N_i} Y_{ij}$$ - total de l'UP n°i

- Tirage aléatoire simple à chaque degré.

$$\hat{T} = \frac{M}{m} \sum_{i \in s} \left(\frac{N_i}{n_i} \sum_{j \in S_i} y_{ij} \right)$$

Remarque: inutile de connaître N pour estimer T.

$$V\left(\hat{T}\right) = \underbrace{M^2 \left(1 - \frac{m}{M}\right) \frac{S_1^2}{m}}_{\text{Variance inter UP}} + \underbrace{\frac{M}{m} \sum N_i^2 \left(1 - \frac{n_i}{N_i}\right) \frac{S_{2,i}^2}{n_i}}_{\text{Variance intra UP}}$$

$$S_1^2 = \frac{1}{M-1} \sum_{i=1}^{M} \left(T_i - \overline{T}\right)^2 \qquad S_{2,i}^2 = \frac{1}{N_i - 1} \sum_{i=1}^{N_i} \left(Y_{ij} - \overline{Y}_i\right)^2$$

SONDAGE A DEUX DEGRÉS

- $S_1^{\,2}$ estimé par $$s_1^2 = \frac{1}{m-1}\sum_{i=1}^{m}\left(\hat{T}_i - \frac{\hat{T}}{M}\right)^2$$ si m>1
- idem pour $S^2_{2,i}$
- Si n_i proportionnel à N_i: taille d'échantillon aléatoire

$$n_i = n_0 \frac{N_i}{N} \qquad E(n_s) = E\left(\sum_{i \in S_i} n_0 \frac{N_i}{N}\right) = \sum_{k \in U_i} n_0 \frac{N_i}{N}\frac{m}{M} = \frac{n_0 m}{M}$$

SONDAGE A DEUX DEGRÉS

- Sondage autopondéré:

  - m unités primaires tirées à probabilités proportionnelles à leur taille
  - tirage d'échantillons de taille fixe n_0

- probabilités d'inclusion constantes

$$\pi_i = \frac{N_j}{N} m \frac{n_0}{N_j} = \frac{mn_0}{N}$$

- Estimateur de la moyenne: N peut être inconnu

$$\hat{\bar{Y}} = \bar{y}$$

SONDAGES A DEUX DEGRÉS

Comment améliorer la précision ?

- Avant tout, construire des UP le plus ressemblantes possible entre elles pour limiter les effets de grappes.
  - Exemple : l'unité ménage est intéressante pour estimer des variables comme le sexe, l'activité, l'âge, etc, mais elle est moins efficace pour étudier le niveau d'instruction, la CS , etc.
- Privilégier le nombre d'UP enquêtées plutôt que le nombre d'US
- Tirer les UP à probabilités inégales
- Stratifier au niveau des UP

CAS PARTICULIER: SONDAGE EN GRAPPES

- Définition: toutes les US sont observées dans les UP tirées.

« Recensement » au deuxième degré

Le tirage systématique est un tirage d'une grappe.

SONDAGE EN GRAPPES

- Cas général :tirage de grappes à probabilités inégales

  - Estimation du total: $$\hat{T} = \sum_{i=1}^{m} \frac{T_i}{\pi_i}$$

  - Estimation d'une moyenne $$\hat{\bar{Y}} = \frac{1}{N} \sum_{i=1}^{m} \frac{N_i \bar{Y}_i}{\pi_i}$$

pb si N inconnu: utiliser l'estimateur de Hajek

$$\widehat{\bar{Y}_{Hajek}} = \frac{\sum_{i=1}^{m} \frac{N_i \bar{Y}_i}{\pi_i}}{\sum_{i=1}^{m} \frac{N_i}{\pi_i}}$$

SONDAGE EN GRAPPES

- Tirage de grappes à probabilités égales

$$\pi_i = \frac{m}{M}$$

➢ taille d'échantillon aléatoire

$$E(n_s) = E(\sum_{i \in S_i} N_i) = \sum_{k \in U_i} N_i \frac{m}{M} = \frac{Nm}{M}$$

➢Estimation

$$\hat{T} = \frac{M}{m} \sum_{i \in s} T_i \qquad V\left(\hat{T}\right) = M^2\left(1 - \frac{m}{M}\right)\frac{S_1^2}{m} \qquad \hat{\bar{Y}} = \frac{M}{m}\frac{1}{N}\sum_{i=1}^{m} N_i \bar{Y}_i$$

SONDAGE EN GRAPPES

- Tirage de grappes à probabilités proportionnelles à la taille

$$\pi_i = m\frac{N_i}{N}$$

$$\hat{\bar{Y}} = \frac{1}{m}\sum_{i=1}^{m}\bar{y}_i$$

$$\hat{V}(\hat{\bar{Y}}) = \frac{1}{m(m-1)}\sum_{i=1}^{m}(1-m\frac{N_i}{N})(\bar{y}_i - \hat{\bar{y}})^2$$

$$E(n_s) = E(\sum_{i\in S_i} N_i) = \sum_{i\in U_i} N_i \frac{N_i m}{N} = \frac{m}{N}\sum_{i=1}^{M} N_i^2$$

SONDAGE EN GRAPPES
conseils pratiques

- Faire des grappes homogènes en inter et hétérogènes en intra (contraire de la stratification).
- Faire beaucoup de grappes de tailles voisines et petites
- En tirer un maximum

Le tirage systématique

- Très utilisé à la place d'un tirage aléatoire à probabilités égales
- Soit N multiple de n. Par exemple on veut tirer 10 individus parmi 1000 : on commence par tirer au hasard un nombre entier entre 1 et 100, si ce nombre est 27, le premier individu sera le n°27, le deuxième le n°127 etc. jusqu'au n°927.
- De façon générale si on a tiré un entier h , les individus sélectionnés ont les numéros : h , $h+M$, $h+2M$, ..., $h+(n-1)M$.
- **Tirage d'une seule grappe** parmi M=N/n grappes.

Le tirage systématique

- L'estimateur de la moyenne est simplement la moyenne de la grappe sélectionnée et sa variance est

$$V\left(\hat{\bar{Y}}\right) = M\sum_{i=1}^{M}\left(\frac{\bar{Y}_i N_i}{N} - \frac{\bar{Y}}{M}\right)^2$$

- Lorsque le fichier se trouve être trié selon un ordre proche de Y, la variance peut être notablement plus faible que pour le tirage aléatoire simple. Exemple $Y_i = i$
- Mais la variance n'est pas estimable .
  - Voir formule $s_1^2 = \frac{1}{m-1}\sum_{i=1}^{m}\left(\hat{T}_i - \frac{\hat{T}}{M}\right)^2$
  - Il est incorrect d'utiliser la variance de l'estimateur du tirage aléatoire simple sauf si la base de sondage a été triée préalablement au hasard.

Tirage systématique: un exemple théorique

$Y_i = i$ Population triée par ordre croissant $N=Kn$

$$\overline{Y} = \frac{N+1}{2} \qquad S^2 = \frac{(N+1)^2}{12}$$

- Tirage équiprobable sans remise :

$$V(\overline{y}_{sr}) = \left(1 - \frac{n}{N}\right)\frac{(N+1)^2}{12n} = \left(1 - \frac{1}{K}\right)\frac{(Kn+1)^2}{12n}$$

- Tirage systématique :

  une grappe : *h, h+K, h+2K,.., h+(n-1)K*

$$\overline{Y}_h = h + \frac{n-1}{2}K$$

- Moyenne

$$E(\overline{Y}_h) = E(h) + \frac{n-1}{2}K = \frac{K+1}{2} + \frac{n-1}{2}K = \frac{nK+1}{2} = \frac{N+1}{2}$$

- Variance

$$V(\hat{\overline{Y}}_{syst}) = V(h + \frac{n-1}{2}K) = V(h) = V(h) = \frac{K^2-1}{12}$$

$$V(\hat{\overline{Y}}_{syst}) < V(\overline{y}_{sr})$$

Exemple *N*=20 *n*=4 $V(\hat{\overline{Y}}_{syst}) = 1.33$ $V(\overline{y}_{sr}) = 7.35$