

Calage sur marges et redressement de la non-réponse

Le cas des enquêtes couplées

Marie Cordier-Villoing (DSCT)
Olivier Sautory (UMS-E)



Introduction

Dispositif d'enquêtes couplées :

Interrogation de plusieurs unités statistiques différentes entre lesquelles existe un lien

Stratégies d'échantillonnages « descendantes » :

Assimilables à un échantillonnage en deux degrés.

Ménages / individus ; Entreprises / salariés

⇒ Différents volets pour une même enquête.

Objectif : comparer les calages indépendants / simultanés et étudier l'impact de la non-réponse sur les calages.

Quelques exemples d'enquêtes couplées

Enquête « COI » : enquête employeurs / salariés

Volet « Entreprises »

Organisation stratégique de l'entreprise

Mobilisation des outils de gestion et d'informatique

Volet « salariés »

Organisation du travail, conditions et rythmes, usage des TIC

Enquête « ECMOSS »

Volet « Employeurs »

Politiques salariales et représentation du personnel

Ou Charges sociales, compléments sociaux et indemnités

Volet « salariés »

Rémunérations et charges

Temps travaillé

Plan

- 1 – Retour sur les calages indépendants
- 2 – Calage simultané en une étape au niveau des unités primaires
- 3 – Calage simultané en une étape au niveau des unités secondaires
- 4 – Calage « simultané » en deux étapes
- 5 – Introduction de la non-réponse au niveau des unités primaires (UP) et unités secondaires (US)

Notations

j : indice d'une unité primaire (UP),

S_I : échantillon des unités primaires j ,

(i, j) : indice d'une unité secondaire (US) appartenant à l'UP j ,

S_j : échantillon des unités secondaires de l'UP j ,

ω : pondérations initiales,

W : pondérations après calage.

1 – Calages indépendants

Principe général

- Les nouvelles pondérations calées des unités échantillonnées sont calculées aussi proches que possible des pondérations initiales,
- L'échantillon aura la même structure que l'ensemble de la population étudiée sur certaines variables.

Les pondérations w_j des UP j vérifient P équations de calage

$$(1) : \forall p \in \{1 \dots P\} : X_p = \sum_{j \in s_I} w_j x_{j,p}$$

Les pondérations $w_{i,j}$ des US (i, j) vérifient S équations de calage

$$(2) : \forall s \in \{1 \dots S\} : X_s = \sum_{j \in s_I} \sum_{i \in s_j} w_{i,j} x_{ij,s}$$

2 – Calage en une étape au niveau des UP

Principe général : Calage réalisé au niveau des UP

- Variables de calage des US remontées au niveau des UP

Estimateur sans biais du total d'une variable de calage pour une UP j :

$$\hat{x}_{j,s} = \sum_{i \in s_j} \omega_{i/j} x_{ij,s} \quad \text{avec les poids conditionnels : } \omega_{i/j} = \frac{\omega_{i,j}}{\omega_j}$$

Les pondérations des UP vérifient P+S équations de calage

$$(1) : \forall p \in \{1 \dots P\} : X_p = \sum_{j \in s_p} w_j x_{j,p}$$

$$(2') : \forall s \in \{1 \dots S\} : X_s = \sum_{j \in s} w_j \hat{x}_{j,s}$$

Par construction, les poids conditionnels des US ne sont pas modifiés :

$$(3) : w_{i,j} = \omega_{i/j} \times w_j \quad \text{i.e.} \quad w_{i/j} = \frac{w_{i,j}}{w_j} = \omega_{i/j}$$

2 – Calage en une étape au niveau des UP

Propriétés cas d'un sondage en grappes

Prop 1 : Établir des statistiques sur les UP à partir du fichier US

- Égalité des poids des UP et des US : $w_{i,j} = w_j$

Prop 2 : Obtenir des estimations sur des nombres d'US identiques à partir des échantillons UP et US

- Ex : pour une enquête ménage interrogeant tous les individus du ménage

Estimation du nombre de femmes de professions intermédiaires dans la population

- $y_{i,j}$ l'indicatrice "être une femme de professions intermédiaires"
- y_j le nombre de femmes de professions intermédiaires du ménage j

$$\hat{Y}_s = \sum_{j \in s_I} \sum_{i \in s_j} w_{i,j} y_{i,j} = \sum_{j \in s_I} w_j \sum_{i \in s_j} \omega_{i/j} y_{i,j} = \sum_{j \in s_I} w_j \sum_{i \in s_j} y_{i,j} = \sum_{j \in s_I} w_j y_j = \hat{Y}_p$$

2 – Calage en une étape au niveau des UP

Propriétés cas d'un sondage en deux degrés

Les deux 1^{ères} propriétés ne sont vérifiées que dans le cas d'un sondage en grappes.

Prop 3 : obtenir des estimations identiques sur des effectifs de regroupements G d'UP à partir des échantillons UP et US

Via l'échantillon des unités primaires : $\hat{N}_{G,p} = \sum_{j \in s_I} w_j N_j \mathbf{I}_{j \in G}$

Via l'échantillon des unités secondaires :

$$\hat{N}_{G,s} = \sum_{j \in s_I} \sum_{i \in s_j} w_{i,j} \mathbf{I}_{j \in G} = \sum_{j \in s_I} \sum_{i \in s_j} w_{i/j} w_j \mathbf{I}_{j \in G} = \sum_{j \in s_I} w_j \mathbf{I}_{j \in G} \sum_{i \in s_j} w_{i/j}$$

- ◆ Cohérence respectée si le tirage au 2^{ème} degré est SAS, SAS stratifié ou équilibré sur la taille des unités primaires

$$\text{i.e. } \sum_{i \in s_j} \omega_{i/j} = \sum_{i \in s_j} w_{i/j} = N_j$$

3 – Calage en une étape au niveau des US

Principe général : Calage réalisé au niveau des US (Esteveao – Särndal)

- Variables de calage des UP intégrées au niveau des US en divisant les variables par l'effectif des UP :

$$\frac{x_{j,p}}{N_j}$$

Les pondérations des US vérifient P+S équations de calage

$$(2) : \forall s \in \{1 \dots S\} : \sum_{j \in s_1} \sum_{i \in s_j} w_{i,j} x_{ij,s} = X_s$$

$$(1') : \forall p \in \{1 \dots P\} : \sum_{j \in s_1} \sum_{i \in s_j} w_{i,j} \frac{x_{j,p}}{N_j} = X_p$$

Les poids des UP sont ensuite calculés tels que :

$$(3') : w_j = \frac{\sum_{i \in s_j} w_{i,j}}{N_j}$$

3 – Calage en une étape au niveau des US

Propriétés

~~Prop 1~~ : Etablir des statistiques sur les UP à partir du fichier US (grappe)

Pas d'égalité des poids des US au sein d'une même UP

~~Prop 2~~ : Estimations sur des nombres d'US identiques (grappe)

$$\hat{Y}_p = \sum_{j \in S_I} w_j y_j = \sum_{j \in S_I} \left(\sum_{i \in S_j} \frac{w_{i,j}}{N_j} \right) \left(\sum_{i \in S_j} y_{i,j} \right) \neq \sum_{j \in S_I} \sum_{i \in S_j} w_{i,j} y_{i,j} = \hat{Y}_s$$

Prop 3 : Estimations identiques sur des effectifs de regroupements G d'UP,

car par construction : $\sum_{i \in S_j} w_{i/j} = \sum_{i \in S_j} \frac{w_{i,j}}{w_j} = N_j$

4 – Calage en deux étapes

Principe général : (Estevao – Särndal)

- Calage sur les UP $\Rightarrow w_j$
- Calage sur les US : les pondérations vérifient S + J équations de calage

$$(2) : \forall s \in \{1 \dots S\} : \sum_{j \in s_I} \sum_{i \in s_j} w_{i,j} x_{i,s} = X_s$$

$$(2'') : \forall j \in s_I : \sum_{i \in s_j} w_{i,j} = N_j * w_j$$

i.e. on impose les contraintes : $\sum_{i \in s_j} w_{i/j} = N_j$

4 – Calage en deux étapes

Propriétés

~~Prop 1~~ : Etablir des statistiques sur les UP à partir du fichier US (grappe)

Pas d'égalité des poids des US au sein d'une même UP

~~Prop 2~~ : Estimations sur des nombres d'US identiques (grappe)

$$\hat{Y}_p = \sum_{j \in S_I} w_j y_j = \sum_{j \in S_I} \left(\sum_{i \in S_j} \frac{w_{i,j}}{N_j} \right) \left(\sum_{i \in S_j} y_{i,j} \right) \neq \sum_{j \in S_I} \sum_{i \in S_j} w_{i,j} y_{i,j} = \hat{Y}_s$$

Prop 3 : Estimations identiques sur des effectifs de regroupements G d'UP

Cf. méthode précédente

5 – Introduction de la non-réponse au niveau des UP et US

Méthodes pour corriger le biais induit par la non-réponse

Méthode en deux temps :

- Correction de la non-réponse
- Calage des poids corrigés de la non-réponse sur les totaux connus dans la population entière

Méthode en un temps :

- Introduction dans le calage des variables explicatives de la non-réponse
- Calage des poids de sondage sur les totaux connus dans la population entière

◆ **On montre que seule la NR au deuxième degré a un impact**

	Prop 1 : (grappes)	Prop 2 : (grappes)	Prop 3 : (2 degrés)
	Stat. sur les UP à partir du fichier US	Estimations sur nombres d'US égales	Estimations sur des effectifs de regroupements G d'UP égales
Calages indépendants	NON	NON	NON
Calages UP en une étape	OUI (CNR 1 étape) NON (CNR 2 étapes)	OUI (CNR 1 étape) NON (CNR 2 étapes)	NON
Calages US en une étape	NON	NON	OUI
Calages en deux étapes	NON	NON	OUI

Calage sur marges et redressement de la non-réponse

Merci de votre attention !

Contacts : Mme Marie Cordier-Villoing
M. Olivier Sautory
Tél. : 01 41 17 65 43
01 41 17 50 82
Courriel : marie.cordier@insee.fr
olivier.sautory@insee.fr



Insee

18 bd Adolphe-Pinard
75675 Paris Cedex 14

www.insee.fr  

Informations statistiques :
www.insee.fr / Contacter l'Insee
09 72 72 4000
(coût d'un appel local)
du lundi au vendredi de 9h00 à 17h00

