

# Vidéo 19 sur 21 : La pondération Calage

---

l'échantillonnage



THE WORLD BANK

MANNHEIM  
BUSINESS SCHOOL

# Qu'est-ce que le calage ? (I)

---

- L'idée générale est d'avoir recours à des variables auxiliaires pour augmenter l'efficacité des estimateurs (Deville et Särndal, 1992)
- **Données auxiliaires** : informations disponibles pour l'ensemble de la base de sondage ou de la population cible, soit pour chaque unité de population soit sous forme agrégée
  - Elles peuvent provenir de la base, de fichiers administratifs, de statistiques publiées ou d'autres sources
  - Enquête auprès des entreprises : la base peut contenir le nombre d'employés d'une période précédente pour chaque société
  - Enquête auprès des ménages : le nombre de personnes dans des groupes définis selon l'âge, l'origine ethnique et le sexe peuvent être publiés à partir d'un recensement ou de projections de population jugées très précises

# Qu'est-ce que le calage ? (II)

---

- Les variables auxiliaires sont généralement incluses dans les question de l'enquête :
  - Veillez à inclure des variables prédictives des résultats de l'enquête
- Les avantages potentiels du calage sont, entre autres :
  - Réduction de la variance d'échantillonnage
  - Correction du biais de couverture de la base et des autres erreurs de la base
  - Correction de la non-réponse
- En général, la plupart des estimateurs standard sont des cas particuliers d'estimateurs calés (par ex. estimateur par expansion, estimateur par le ratio, estimateur post-stratifié, estimateur GREG, estimateur par ratissage)
- La plupart des outils statistiques proposent les principales fonctions d'ajustement par calage

# Post-stratification (I)

---

- Particulièrement fréquente dans les enquêtes de ménages où les variables auxiliaires sont des indicateurs de groupes démographiques (catégorie d'âge, sexe et origine ethnique)
- Mise en œuvre avec des classes de pondération par calage créées en croisant toutes les catégories de variables qualitatives et en construisant une pondération reproduisant les effectifs de population par classe dans les estimations pondérées
- Bon moyen d'exploiter les variables auxiliaires qui vous semblent être de bons indicateurs de variables importantes, qui ont été collectées dans le cadre de l'enquête mais ne peuvent pas être aisément utilisées pour la sélection d'échantillon

# Post-stratification (II)

---

- Par exemple, post-stratification par le sexe et le niveau d'éducation :
  - Sexe (1 = Homme, 2 = Femme)
  - Niveau d'éducation (1 = collège ; 2 = baccalauréat ; 3 = enseignement supérieur ; et 4 = licence ou plus)
- La classification croisée aboutit à  $G = 8$  classes pouvant être utilisées pour la post-stratification
- L'emploi de nombreuses variables auxiliaires importantes pour la post-stratification permet de réduire le biais
- Cependant, cela peut entraîner des classes de pondération vides ou avec très peu de cas de répondants
  - Entraîne des estimations instables des contrôles de la population et ajoute une variabilité inutile à la pondération finale

# Post-stratification (III)

---

- L'ajustement de post-stratification se calcule comme suit

$$a_{g,i} = \frac{N_g}{\hat{N}_g}$$

où  $N_g$  est l'effectif de population de la post-strate  $g$

et  $\hat{N}_g = \sum_{s_g} d_i$  son estimation à partir de la pondération d'entrée  $d_i$

- On calcule alors le poids post-stratifié final pour l'unité  $i$  de la post-strate  $g$  de la manière suivante :

$$w_i = d_i \times a_{g,i}$$

# Post-stratification (IV)

## Exemple

---

Sexe	Âge	$N_g$	$\hat{N}_g$	$a_{g,i}$
Homme	18-34	16 516 047	13 138 357	1,25708
Homme	35-44	9 722 899	6 149 134	1,58118
Homme	45-54	10 928 639	9 911 012	1,10267
Homme	55-64	8 488 197	14 837 247	0,57208
Homme	≥65	8 540 787	12 204 434	0,69981
Femme	18-34	17 200 469	9 210 347	1,86751
Femme	35-44	10 343 299	14 061 085	0,73559
Femme	45-54	11 485 779	15 893 681	0,72266
Femme	55-64	9 334 197	18 086 949	0,51607
Femme	≥65	11 055 919	25 539 782	0,43289

# Ratissage (I)

---

- Le ratissage est une procédure d'ajustement dans laquelle les estimations sont contrôlées par les totaux marginaux de la population
  - Présuppose implicitement que les interactions entre les variables de calage ne sont pas importantes pour expliquer les variables de l'enquête
- Le principal avantage du ratissage par rapport à la post-stratification est que le ratissage peut permettre l'emploi d'un plus grand nombre d'informations auxiliaires (seuls les totaux marginaux sont nécessaires, pas besoin des totaux de catégories croisées)
- Réduit également les problèmes de cellules de pondération vides ou quasiment vides
- Dans l'exemple suivant où le ratissage s'applique, seuls les effectifs marginaux du sexe et de l'âge sont nécessaires



# Ratissage (II)

## Exemple (version 1)

---

Sexe	$N_g$	$\hat{N}_g$	$a_{g1,i} = N_g / \hat{N}_g$
Homme	54 196 569	56 240 183	0,96366
Femme	59 419 663	82 791 843	0,71769

- On actualise le poids en calculant  $d_i^{(1)} = d_i \times a_{g1,i}$
- On emploie les poids actualisés pour calculer  $\hat{N}_g$  ci-dessous

Âge	$N_g$	$\hat{N}_g$	$a_{g2,i} = N_g / \hat{N}_g$
18-34	33 716 516	19 271 206	1,74957
35-44	20 066 198	16 017 324	1,25278
45-54	22 414 418	20 957 760	1,06950
55-64	17 822 394	27 279 096	0,65333
$\geq 65$	19 596 706	30 090 847	0,65125

- On actualise le poids en calculant  $d_i^{(2)} = d_i^{(1)} \times a_{g2,i}$
- On recommence le processus jusqu'à ce que les poids convergent

---

*FIN DE LA VIDÉO 19*