

Bases de sondages multiples et redressement des poids extrêmes

Le cas de l'enquête auprès des
intervenants au domicile de
personnes fragilisées (IAD – 2008)

Plan de la présentation

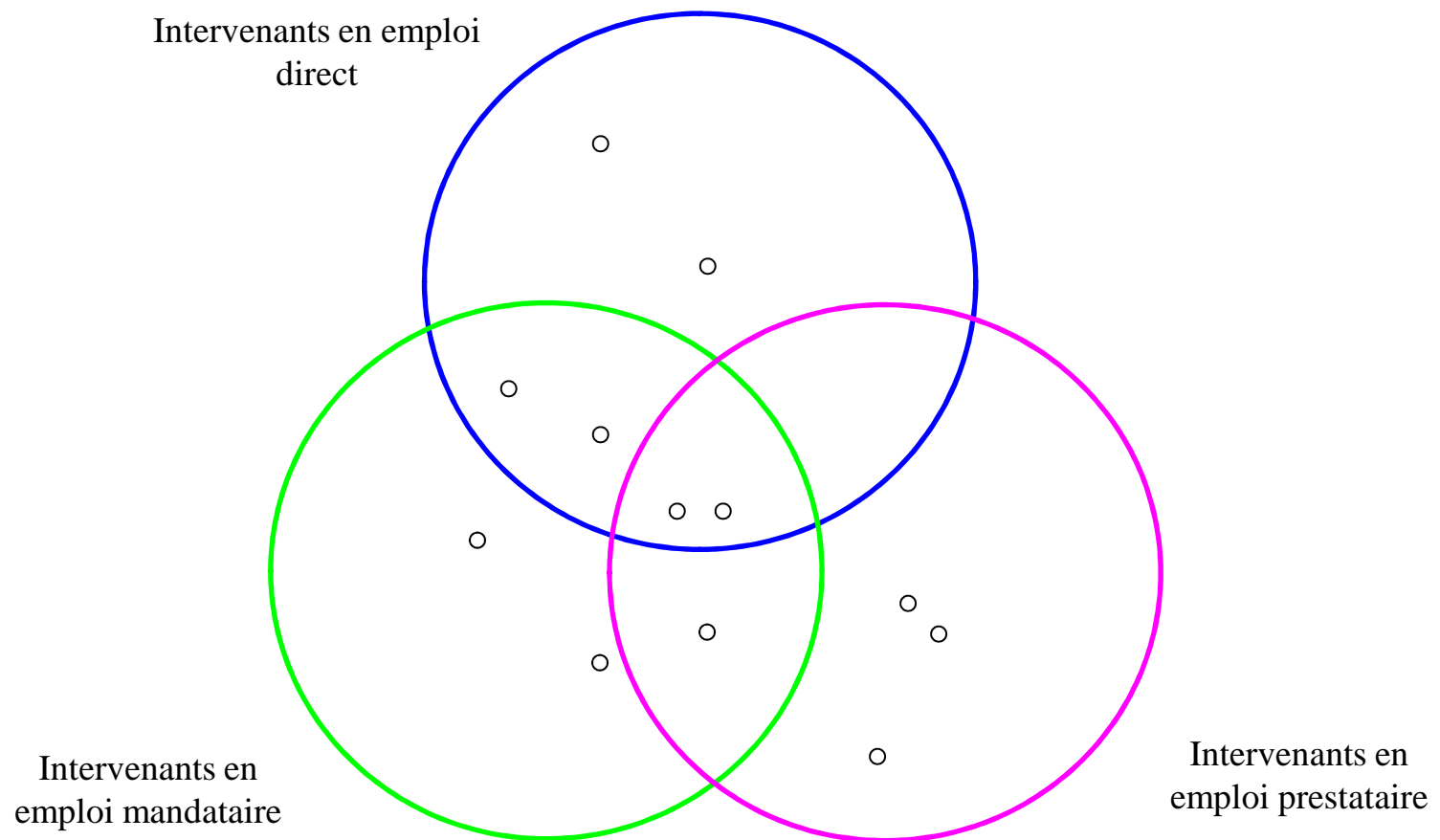
- Protocole d'enquête
- Le partage des poids
- Les premières pondérations
- Technique de troncature des poids extrêmes

Protocole d'enquête

Champ de l'enquête

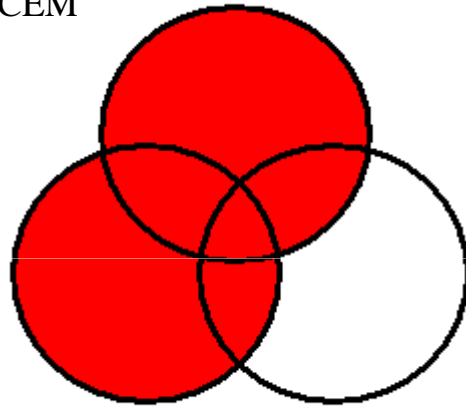
- “ Intervenants au domicile de personnes âgées, handicapées, ou de toute autre personne nécessitant de l'aide de façon temporaire ou durable dans les actes essentiels de la vie quotidienne ”
- Trois modes d'exercice possibles :
 - Prestataire
 - Mandataire
 - Emploi direct

Sélection des intervenants (1)

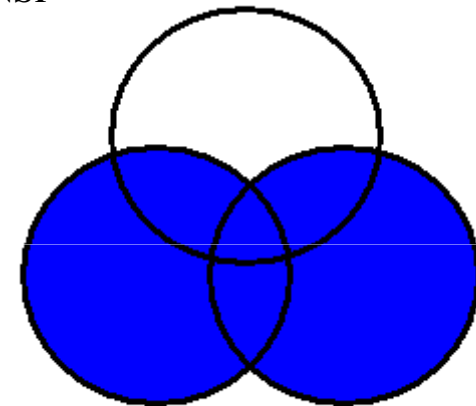


Sélection des intervenants (2)

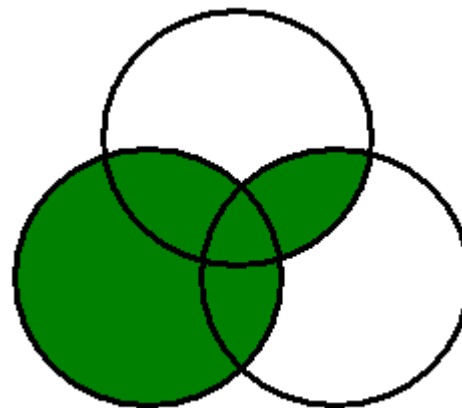
Base IRCEM



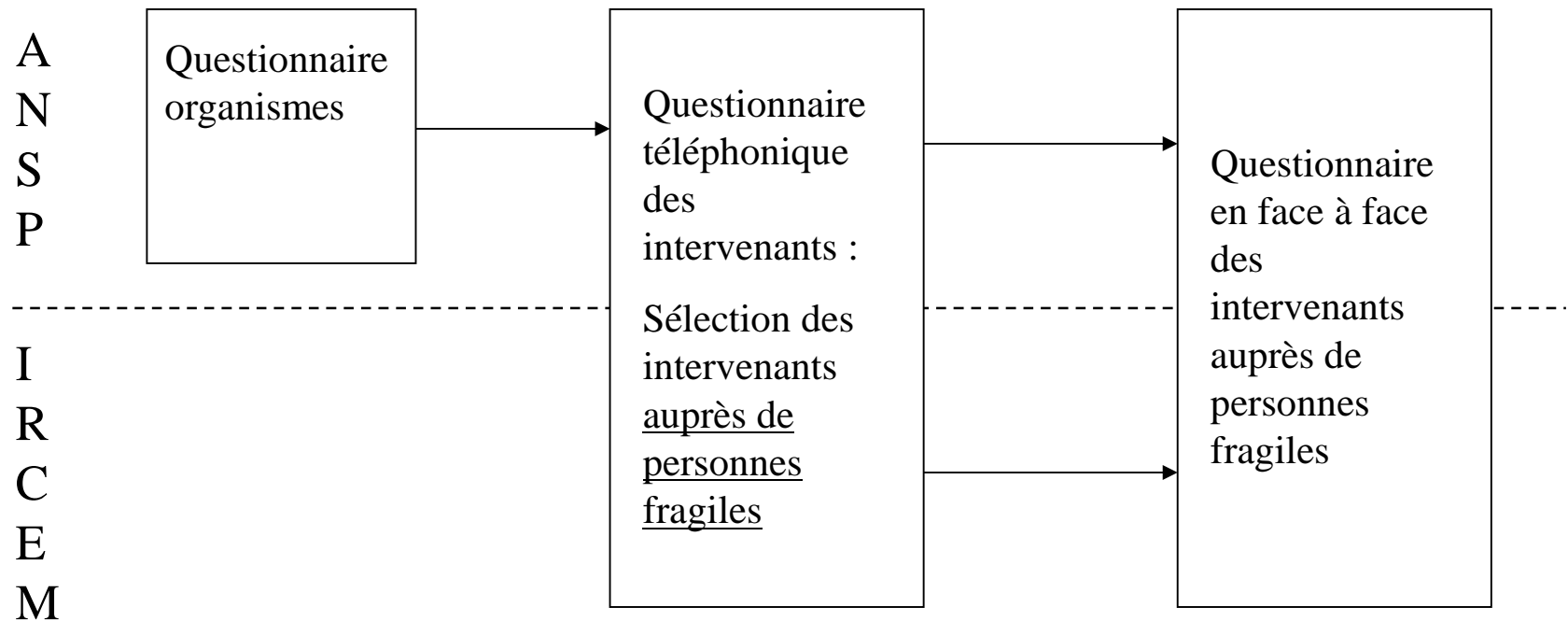
Base ANSP



Couverture
commune



Sélection des intervenants (3) passation des questionnaires



Le partage des poids

Principe

- Constat : on ne connaît pas le recouvrement des bases de sondage
- Principe : compter les liens entre les deux bases
 - Pour chaque individu tiré
 - Au moment de l'interview
- Diviser le poids total de chaque individu par le nombre de fois où il aurait pu être tiré

Dans notre cas...

- Intervenants en emploi direct exclusif
 - Base IRCCEM uniquement
 - Pas de doublon ($L_i=1$)
- Intervenants en emploi prestataire exclusif
 - Base ANSP uniquement
 - Doublons possibles (si plusieurs organismes) [$L_i \geq 1$]
- Intervenants en mode mandataire
 - Bases IRCCEM ou ANSP
 - Pas de doublon dans la base IRCCEM mais doublons possibles dans la base ANSP (si plusieurs organismes)
 - $L_i \geq 2$

De façon un peu plus formelle

$$L_i = \underbrace{1_{\{i \in \Omega_d\}} + 1_{\{i \notin \Omega_d\}} 1_{\{i \in \Omega_m\}}}_{\text{Liens IRCEM}} + \underbrace{M_i + P_i}_{\text{Liens ANSP}}$$

Les premières pondérations

Les intervenants de la base IRCEM

- Tirage à deux degrés et stratifié
 - Départements (PPT)
 - Intervenants (SAS stratifié)
- Redressements par deux calages successifs
 - L'interview téléphonique (taux de réponse : 30%)
 - L'interview en face à face (taux de réponse : 79%)

Les intervenants de la base ANSP

- Tirage à 3 degrés stratifié
 - Département (PPT)
 - Organismes (PPT et SAS stratifiés)
 - Intervenants (Systématique)
- Redressements par trois calages successifs
 - Les organismes (taux de réponse : 30%)
 - L'interview téléphonique (taux de réponse : 77%)
 - L'interview en face à face (taux de réponse : 82%)

Ça se corse...

Statistiques sur les poids initiaux des intervenants des organismes

Moyenne	357,8
Médiane	121,3
Écart-type	1 180,0
Rapport Q3/Q1	4,1
Rapport C95/C5	76,7
Rapport Max/Min	2 699,0

Ouïe...

Le cas des poids extrêmes : que s'est-il passé ?

Le cas d'école (1)

- Tirage du 1^{er} degré : $\pi_d = 30 \cdot \frac{X_d}{\sum_{d=1}^{96} X_d}$
- Tirage du 2nd degré : $\pi_{dm} = 29 \cdot \frac{X_{dm}}{\sum_{m=1}^{M_d} X_{dm}}$
- Tirage du 3^{ème} degré : $\pi_{dmi} = \frac{n}{X_{dm}} \quad \forall i$

Le cas d'école (2)

- Au final :

$$\pi_i = \pi_d \cdot \pi_{dm} \cdot \pi_{dmi} = 30 \cdot \frac{X_d}{\sum_d X_d} \cdot 29 \cdot \frac{X_{dm}}{\sum_m X_{dm}} \cdot \frac{n}{X_{dm}}$$

$$\pi_i = 870 \cdot \frac{n}{\sum_d X_d} \quad \forall d, m, i$$

- La dispersion ne vient que du redressement de la non réponse totale

Ce qu'il s'est passé : information incomplète...

- La base de sondage ne contenait que des effectifs salariés des organismes (*i.e.* en mode prestataire)
- ... ce qui pose problème, sauf dans des cas très particuliers

En effet, si l'on pose : $X_{dm}^M = f(X_{dm}^P) + \varepsilon_{dm}$

$$\pi'_i = 870 \cdot \frac{X_{dm}^P}{\sum_d X_d^P} \cdot \frac{n}{X_{dm}^P + X_{dm}^M} = 870 \cdot \frac{X_{dm}^P}{\sum_d X_d^P} \cdot \frac{n}{X_{dm}^P + f(X_{dm}^P) + \varepsilon_{dm}}$$

$$\pi'_i = A \cdot \frac{X_{dm}^P}{X_{dm}^P + f(X_{dm}^P) + \varepsilon_{dm}}$$

... et imparfaite

- Les effectifs salariés pouvaient dater de plusieurs années...
- ... dans un contexte institutionnel particulièrement mouvant :
 - 1,18 millions de salariés dans le secteur des services à la personne en 2006
 - 1,28 millions en 2007
 - 1,32 millions en 2008

Régler le problème des poids extrêmes

- Idée : tronquer les poids les plus élevés $w_i \rightarrow w_{it}$
- Les contraintes :
 - La taille de la population estimée doit rester la même

$$\sum_{i \in S} w_{it} = \sum_{i \in S} w_i$$

- Minimiser les écarts d'estimation par rapport à la pondération initiale (sans biais)
 - Analyse des écarts absolu et relatif des estimateurs
 - Analyse de l'erreur quadratique moyenne

$$EQM = (\hat{Y}_t - \hat{Y})^2 + V(\hat{Y}_t)$$

La solution de Potter (1990)

$$w_{it} = \tau_i w_0 + (1 - \tau_i) w_i \frac{\sum_{i \in S} (1 - \tau_i) \frac{w_0}{w_i}}{\sum_{i \in S} (1 - \tau_i) w_i}$$

avec $\tau_i = 1$ si $w_i \geq w_0$, 0 sinon

- Problème : 1^{ère} contrainte non satisfaite $\sum_{i \in S} w_{it} \neq \sum_{i \in S} w_i$

Garder le principe

1. on garde la forme générale

$$w_{it} = \tau_i w_0 + (1 - \tau_i) w_i A$$

2. en incluant la contrainte $\sum_{i \in S} w_{it} = \sum_{i \in S} w_i$

3. soit $\sum_{i \in S} [\tau_i w_0 + (1 - \tau_i) w_i A] = \sum_{i \in S} w_i$

Au final

$$w_{it} = \tau_i w_0 + (1 - \tau_i) w_i \frac{\sum_{i \in S} (w_i - \tau_i w_0)}{\sum_{i \in S} (1 - \tau_i) w_i}$$

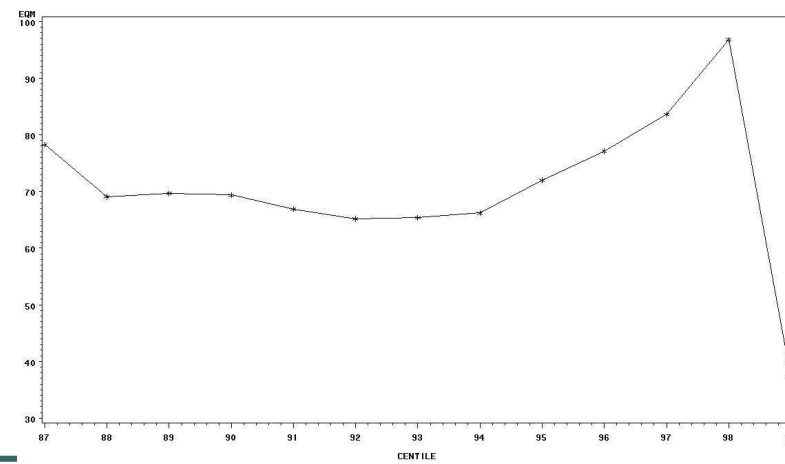
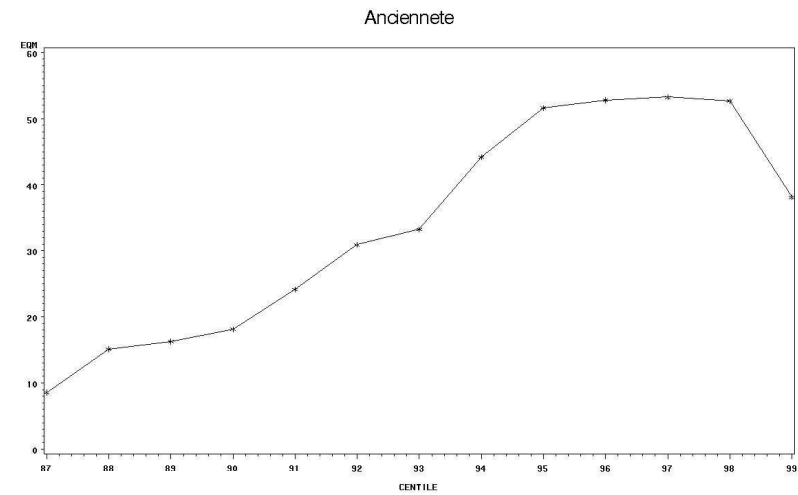
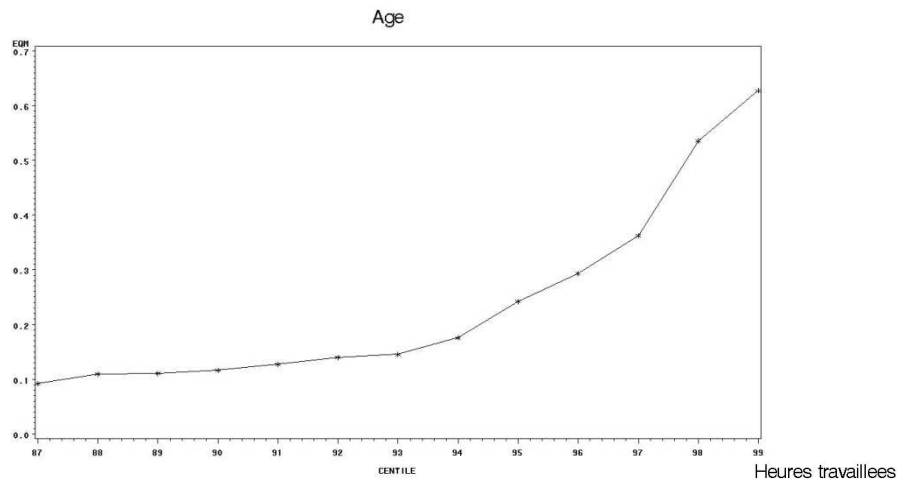
avec $\tau_i = 1$ si $w_i \geq w_0$, 0 sinon

- Problème : on a des poids supérieurs à w_0
- Solution :
 - recommencer la troncature en remplaçant w_i par w_{it}
 - réitérer l'opération jusqu'à vérifier $Max(w_{it}) = w_0$

Intégrer la troncature dans la pondération

1. Redresser l'échantillon par calage
2. Tronquer les poids
 1. À partir des centiles des poids initiaux dans l'enquête IAD
 2. Regarder l'évolution de l'EQM et des biais relatif et absolu en fonction de la troncature considérée
3. Re-caler l'échantillon en fonction :
 1. Des premières variables de calage
 2. Des variables d'intérêt (en prenant comme estimateurs sans biais ceux calculés avec la pondération issue du premier calage)

Exemple d'application (1)



Exemple d'application (2)

Statistiques sur les poids initiaux des intervenants des organismes

Moyenne	357,8
Médiane	121,3
Écart-type	1 180,0
Rapport Q3/Q1	4,1
Rapport C95/C5	76,7
Rapport Max/Min	2 699,0

Statistiques sur les poids tronqués et calés des intervenants des organismes (téléphone)

Moyenne	464,6
Médiane	389,3
Écart-type	395,2
Rapport Q3/Q1	1,9
Rapport C95/C5	25,4
Rapport Max/Min	101,8