
Méthodes probabilistes pour enquêter en l'absence de base de sondage

Maryse Marpsat (INSEE-INED-ERIS)

Nicolas Razafindratsima (INED)

Rencontres internationales sur les

Enquêtes auprès des populations difficiles à joindre

INED – 19 mai 2009



Objectifs de la présentation

- Caractériser les ‘populations difficiles à joindre’
 - Présenter quelques méthodes permettant de tirer un échantillon ‘extrapolable’, i.e. :
 - ❑ Soit dont la probabilité de sélection est calculable (échantillon probabiliste)
 - ❑ Soit à partir duquel on peut produire une estimation sans biais de certaines variables sur toute la population d’intérêt
-

Principales difficultés conduisant à classer une population comme « difficile à joindre »

- Effectif relativement faible
 - Identification difficile
 - Pas de base de sondage
 - Les personnes ne souhaitent pas révéler qu'elles font partie de la population cible
 - Les comportements de la population d'intérêt sont méconnus
-

Méthodes probabilistes pour enquêter en l'absence de base de sondage

- Techniques « classiques » :
 - Le sondage en plusieurs degrés (ou en grappes).
Condition : avoir une base d'unités primaires
 - L'enquête en deux phases
 - Inconvénients : coût élevé, ne s'appliquent que pour des populations assez stables, facilement localisables, etc.
-

Méthodes probabilistes pour enquêter en l'absence de base de sondage

- Il existe des techniques plus adaptées aux populations difficiles à joindre
 - Principe : trouver où les personnes sont présentes (dans des lieux précis, dans le réseau de relations de quelqu'un), les enquêter, et produire un estimateur sans biais à partir de certaines questions
-

Time-location sampling (TLS)

- Principe : dresser la liste exhaustive des lieux fréquentés par les personnes, avec les moments où ils sont fréquentés
 - Tirage de lieux * moments
 - Tirage de personnes dans les lieux sélectionnés aux moments sélectionnés
 - Calcul d'une pondération (Deville, 1999 ; Lavallée, 1995, 2002, 2007)
-

Time-location sampling (TLS)

- Problèmes :
 - ❑ l'établissement et l'actualisation de la liste (long, cher, informations recueillies pas toujours fiables – notamment sur les files actives)
 - ❑ les personnes appartenant à la population d'intérêt mais qui n'utilisent pas ou très peu ces lieux (si elles sont en forte proportion on a un biais)
 - ❑ le calcul de la pondération (par ex. par la méthode de partage des poids) demande l'inclusion de questions spécifiques
 - ❑ les lieux en question peuvent poser des problèmes de collecte : refus des gestionnaires/patrons etc., départ rapide des utilisateurs/ clients...
-

Respondent-driven sampling (RDS)

- Proche de la méthode « boule de neige », mais possibilité de calculer un estimateur non biaisé (Heckathorn, 1997)
 - Principe :
 - On part de personnes appartenant à la population cible
 - Ils reçoivent un nombre limité de coupons (3 ou 4) qu'ils utilisent pour recruter d'autres personnes
 - si une personne est recrutée, le recruteur est rémunéré
 - La personne recrutée reçoit un nombre de coupons identique et recrute à son tour
 -
 - On s'arrête quand la taille de l'échantillon visée est obtenue
 - Calcul d'une pondération
-

Respondent-driven sampling (RDS)

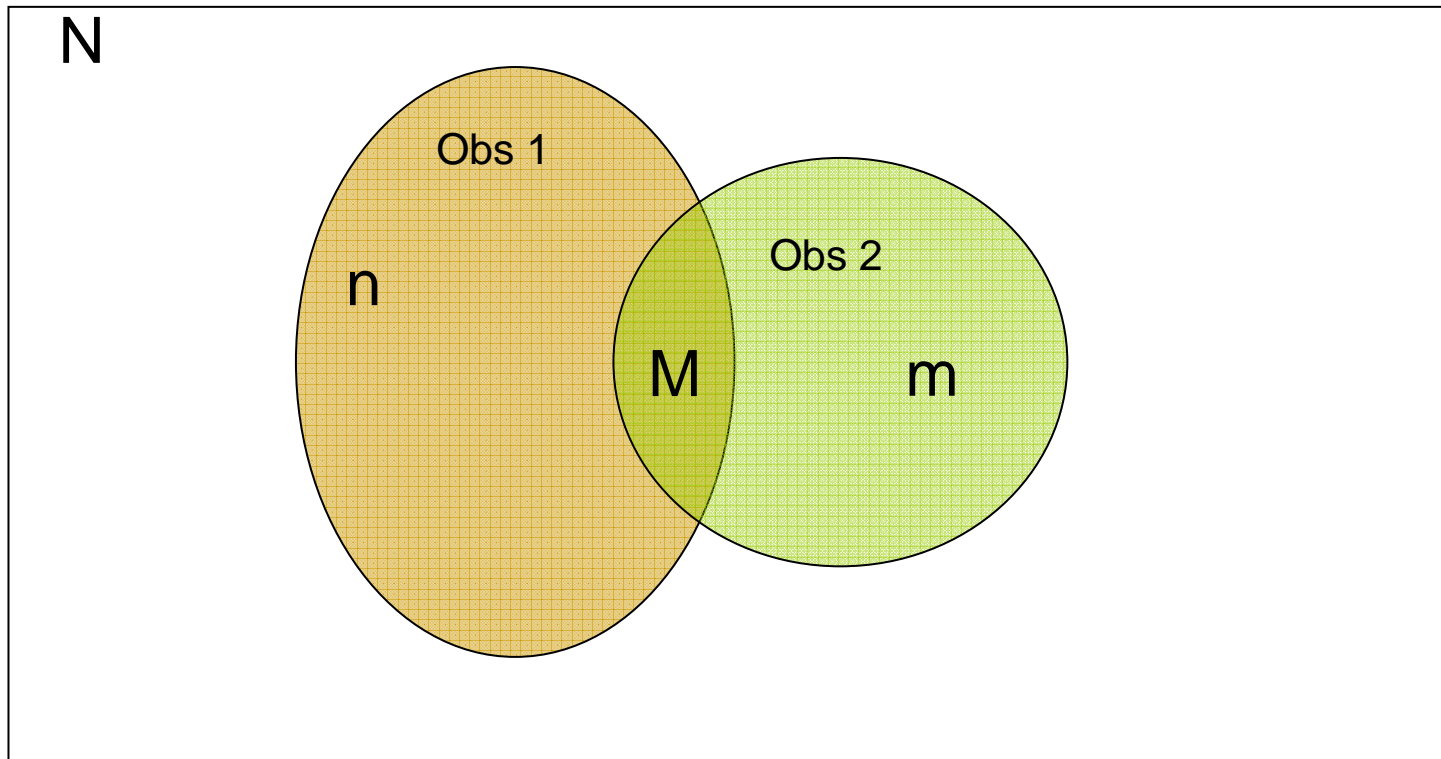
■ Problèmes :

- ❑ Faible probabilité de sélection pour les personnes à faible réseau social
 - ❑ La taille du réseau du recruteur n'est pas évidente à connaître
 - ❑ Il est nécessaire que les personnes se connaissent comme faisant partie de la population d'intérêt
 - ❑ Organisation de la collecte, de la rémunération, etc.
-

Capture-recapture (ou contact-recontact)

- 2 observations (ou sources) indépendantes de la population
- N =taille de la population d'intérêt, n =nombre observé la 1^{ère} fois, m nombre observé la 2^e fois, M nombre observé les 2 fois :
$$N=(m*n)/M$$
- Suppose de pouvoir repérer les mêmes personnes d'une observation à l'autre (besoin d'un identifiant quelconque)

Capture-recapture (ou contact-recontact)



$$N = (n * m) / M$$

Capture-recapture

- Problèmes : plusieurs hypothèses doivent être remplies pour que le modèle soit valable
 - Les individus doivent avoir la même probabilité d'être observés à chaque période
 - Les observations à 2 périodes doivent être indépendantes
 - La population ne doit pas se modifier
 - Une modélisation log-linéaire permet de s'affranchir de ces contraintes, mais les informations à introduire ne sont pas aisées à établir
-

Time Location Sampling

- *Échantillonnage indirect et méthode généralisée de partage des poids : application au domaine du tourisme* - Jean-Claude DEVILLE, Laboratoire de statistique d'enquêtes, Ensaï/Crest.
 - *Conception, échantillonnage et organisation du terrain d'une enquête TLS (1) : le cas des sans domicile* - Martine QUAGLIA et Géraldine VIVIER, Ined.
 - *Conception, échantillonnage, organisation du terrain d'une enquête TLS et nouveaux développements (2) : Coquelicot, une enquête auprès d'usagers de drogues* - Marie JAUFFRET, Yann LE STRAT, InVS.
-

Respondent Driven Sampling et Capture-recapture

- *Respondent Driven Sampling: Lessons and challenges from a decade of sampling hard to reach populations* - Keith SABIN, OMS, Genève.
 - *Can we measure homelessness? A critical evaluation of the method of « capture-recapture »* - Malcolm WILLIAMS, University of Plymouth.
-

Comparaison des méthodes

- *Opportunities and Challenges of Time-Space Sampling and Respondent-Driven Sampling of Hard-to-Reach Populations* - Salaam SEMAAN, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, Centers for Disease Control and Prevention – Atlanta.
 - *Enquêtes auprès des usagers de drogue : l'expérience de l'OFDT* - Agnès Cadet-Tairou, Observatoire Français des Drogues et des Toxicomanies - Paris.
-