Méthodologie de l'Enquête sociale générale Cycle 20 (2006) Les transitions familiales

Présentation au centre interuniversitaire québécois
de statistiques sociales

Isabelle Marchand
Méthodologiste principale
Statistique Canada
5 octobre 2007

Aperçu de la présentation

- Méthodologie
- L'Enquête sociale générale (ESG)
- Plan de sondage de l'ESG-20
- Questions?

Méthodologie

- Besoin d'information
- Enquête
 - Activité organisée et méthodique de collecte de données sur des caractéristiques d'intérêt d'une partie ou de la totalité des unités d'une population à l'aide de concepts, de méthodes et de procédures bien définis. Elle est suivie d'un exercice de compilation permettant de présenter les données recueillies sous forme récapitulative utile
- Méthodologie
 - Théorie statistique + théorie de l'échantillonnage et des méthodes d'enquêtes + connaissance de la pratique

Méthodologie

■ Échantillonnage

- Moyen de **sélection d'un sous-ensemble d'unités** d'une population, dans le but de **recueillir** des données pour ces unités, afin de **produire de l'inférence** sur la population visée

■ Deux types d'échantillonnage

1. Probabiliste: Organisme national statistique, Statistique Canada (STC)
2. Non-probabiliste: le lien à la population suppose que l'échantillon est directement représentatif -> pas à STC

L'Enquête sociale générale (ESG)

- Établie en 1985 pour combler certaines lacunes parmi les statistique sociales de Statistique Canada (STC)
- Rassembler des données sur les tendances sociales de manière à suivre l'évolution du mode de vie des canadiens
- Fournir des renseignements immédiats sur des questions de politiques sociales précises

L'Enquête sociale générale (ESG)

- Besoins et contraintes de l'ESG
 - Analyse détaillée au niveau national sur la population canadienne adulte
 - Comparabilité provinciale, urbaine et rurale
 - Coût acceptable
 - À buts multiples
 - Fichier de micro-données à grande diffusion
 - Affectent le choix du plan de sondage

L'Enquête sociale générale (ESG)

- Enquête nationale annuelle à buts multiples
- Thème cyclique – 5 ans
- Contenu commun d'un cycle à l'autre (tendance)
Contenu spécifique (immédiat)
- Mode de collecte : téléphone
- Recueillies auprès de 25 000 répondants depuis ESG-13 (avant 10 000)
- ESG-20 fait suite à l'ESG-15: Enquête rétrospective sur la famille

L'Enquête sociale générale (ESG)

Les différents thèmes

Thèmes	Année de référence (# cycle)			
	1er	2e	3e	4e
Santé	1985 (1)	1991 (6)		
Emploi du temps	1986 (2)	1992 (7)	1998 (12)	2005 (19)
Victimisation	1988 (3)	1993 (8)	1999 (13)	2004 (18)
Éducation, travail, retraite	1989 (4)	1994 (9)	2002 (16)	2007 (21)
Famille	1990 (5)	1995(10)	2001 (15)	2006 (20)
Soutien social et vieillissement	1985 (1)	1990 (5)	1996 (11)	2002 (16) 2007 (21)
Accès et utilisation d'internet	2000(14)			
Engagement social	2003(17)	2008 (22)		

Plan de sondage de l'ESG-20

Activités pré-collecte

- Population cible / population d'enquête
- Base de sondage
- Répartition de l'échantillon
- Sélection de l'échantillon
- Questionnaire

Plan de sondage de l'ESG-20

Population cible, population d'enquête

- Population cible (veut estimer!)
 - Personnes de 15 ans et plus dans les 10 provinces canadiennes, excluant les pensionnaires d'établissements
- Population d'enquête (peut collecter!)
 - Personne appartenant à un ménage avec une ligne résidentielle terrestre
- La différence entre les 2 doit être minimale!

Plan de sondage de l'ESG-20

Base de sondage

- Outil qui permet
 - d'identifier (probabilité de sélection)
 - classifier (national, provincial, urbain, rural)
 - rejoindre la population d'enquête (téléphone)
- La sélection de l'échantillon se fera à partir de la base de sondage
- Information clé: poids initial de chaque unité – inverse de la probabilité de sélection sur la base

Plan de sondage de l'ESG-20

Base de sondage

- ESG-20: « Méthode de composition aléatoire avec élimination de banques inactives »
 - Utilisation de grappes (banques) de 100 numéros contenant au moins un # de téléphone résidentiel
 - Ces grappes peuvent être attribuées à un secteur géographique (limite du détail!)
 - Génération d'un numéro de téléphone au hasard parmi les grappes qui seront choisies (nouveaux numéros et numéros confidentiels)
 - Bon compromis entre coût et qualité

Plan de sondage de l'ESG-20

Répartition de l'échantillon

- *Où attribuer les 25 000 répondants pour être le plus efficace possible?*
- Stratification: partition de la population à des fins d'échantillonnage
- ESG-20 - Chaque province, selon l'importance de la taille de la population, est divisée en
 - Région(s) métropolitaine de recensement (RMR)
 - Regroupement des autres RMR
 - Non-RMR
- Les 10 provinces sont divisées en 27 strates

Plan de sondage de l'ESG-20

Répartition de l'échantillon

- 25 000 répondants répartis à l'intérieur de chaque strate de telle façon à permettre une bonne analyse
 - Nationale
 - Comparable entre provinces
 - Comparable entre urbain et rural
 - Certaines estimations générales aient une variabilité suffisamment faibles
- 2 méthodes: « Kish » et « min_p »
- Chaque taille ainsi obtenue est gonflée pour tenir compte de la non-réponse et du fait que ce sont des numéros de téléphone qui sont échantillonnés (base de sondage!)

Plan de sondage de l'ESG-20

Sélection de l'échantillon

- Sélection aléatoire simple avec remise d'une grappe de 100 numéros de téléphones dans chaque strate
- Enquête annuelle fractionnée en sélection mensuelle indépendante (saisonnalité)

Plan de sondage de l'ESG-20

Questionnaire

- Contenu commun d'un cycle à l'autre (tendance)
 - Ex: Mariages et unions libres
- Contenu spécifique
 - Ex: Équilibre entre vie professionnelle et la vie familiale
- Variables de classification
 - Ex: Âge, sexe, langue, occupation
- Variables nécessaires à la pondération
 - Ex: Nombres de lignes téléphoniques, code postal

Plan de sondage de l'ESG-20

Collecte

- Données recueillies par téléphone de juin 2006 à octobre 2006
- Collecte ITAO – interview téléphonique assisté par ordinateur
- Distribution uniforme de l'échantillon sur 4 périodes de collecte de 2 mois pour neutraliser l'effet saisonnier
- Pré-collecte par une compagnie externe privée (enlève les numéros hors-service)

Plan de sondage de l'ESG-20

Collecte

- 4 bureaux régionaux : Halifax, Sherbrooke, Winnipeg et Edmonton
- Pas d'entrevue par procuration
- Durée de l'entrevue : environ 45 minutes
- 23 608 répondants, 35 045 ménages
- Taux de réponse : 67%

Plan de sondage de l'ESG-20

Traitement

Pour qu'une question se transforme en information au niveau de la population -> besoin de traitement et de pondération!!

- Vérification (différente de celles de l'ITAO)
- Variables dérivées
- Codage
- Géocodage en utilisant le code postal

Plan de sondage de l'ESG-20

Pondération

- La dérivation du poids personne est complexe et nécessite plusieurs ajustements au poids initial
 - # de téléphone -> ménages résidentiels individuels
 - Non-réponse au niveau ménage
 - Ménage -> personne
 - Non-réponse au niveau personne
 - Période de collecte mensuelle -> information annuelle
 - Population d'enquête -> population cible

Plan de sondage de l'ESG-20

Pondération

- Poids personne
 - Enquête au niveau personne
 - Plusieurs ajustements
 - À la fin du processus de pondération, les totaux de population au niveau province x sexe x groupes d'âge correspondent aux projections du recensement de 2001 pour la période de référence 2006.
- Poids ménage disponible
 - N'est pas soumis au même traitement

Plan de sondage de l'ESG-20

Pondération

- Erreur d'échantillonnage (échantillonnage et non un recensement)
 - Variance, coefficient de variation (CV), intervalle de confiance (écarts-types)
- Estimée à l'aide de la « méthode du bootstrap moyen »
 - Plan complexe -> pas de variance exacte mais basée sur des techniques de répliques
 - 500 poids bootstrap disponibles
 - Moyen – permet d'avoir des poids bootstrap sur le fichier de micro-données (facteur – 25)

Plan de sondage de l'ESG-20

Analyse

- Les répondants ne forment pas un échantillon aléatoire simple de la population.
- Plan complexe, stratifié, à plusieurs degrés, avec probabilités inégales de sélection
- Analyse -> utilisation des poids personnes et des 500 poids bootstrap

Plan de sondage de l'ESG-20

Analyse

- Analyse descriptive ou analytique
 - Calcul de la précision:
 - Utilisation de logiciel avec les poids bootstrap et facteur de correction du bootstrap moyen
 - Diffusion d'estimations avec une trop grande variabilité – lignes directrices (CV)
 - 0 à 16,5% - aucune restriction
 - 16,6% à 33,3% - mise en garde
 - 33,4% - non recommandée, sinon avec grande prudence

Questions ?

- Merci!