



Statistique
Canada

Statistics
Canada

Canada



Statistique Canada
www.statcan.gc.ca



Stratégie de sélection et de coordination d'échantillons pour les enquêtes à un et deux degrés

Jean-Louis Tambay
**XI^{èmes} Journées de Méthodologie
Statistique de l'INSEE**
24-26 janvier 2012



Plan de l'exposé

- Plan de sondage de l'Enquête sur la population active (EPA)
- Caractéristiques souhaitables d'un système
- Stratégie de base
- Quelques résultats
- Coordination de la collecte
- Reproduction du plan de sondage de l'EPA
- Alternative à l'utilisation de départs
- Chevauchement avec les enquêtes à 1 degré
- Prise en considération des tendances
- Conclusion

Plan de sondage à 2 degrés de l'EPA

- Méthode Rao-Hartley-Cochran dans chaque strate
 - **Grappes** réparties parmi (six) groupes de renouvellement
 - Séquençage aléatoire des grappes et de leurs départs
 - **Départs** = échantillons systématiques de logements
 - Ex. 4 grappes, 33 départs d'un groupe de renouvellement

N° grappe	219 (9 départs)	082 (7dép.)	447 (8 dép.)	153 (9 dép.)
N° départ	6 7 8 9 1 2 3 4 5	4 5 <u>6</u> 7 1 2 3	1 2 3 4 5 6 7 8	8 9 1 2 3 4 5 6 7

- Sélection d'un départ par groupe (sél. PPT de grappes)
- Rotation dans 1 groupe à chaque mois – suivre la liste
- Utilisation du plan de l'EPA par d'autres enquêtes
 - Prendre départ(s) suivant(s) dans 1 à 6 des groupes

Caractéristiques du plan de l'EPA

- Minimisation de la rotation des grappes, partage des grappes et collecte coordonnée des enquêtes
 - Réduction des coûts de listage et de collecte
- Gestion de la croissance (nouveaux logements)
 - Alloués séquentiellement aux départs existants
 - **Stabilisation** (sous-éch.) maintient la charge de travail
 - Croissance élevée → sous-éch. dans la grappe même
 - Cas extrêmes → restratification (avec chevauchement)
- Lacunes principales
 - Pas convivial ou optimal pour les autres enquêtes
 - Impact de croissance élevée/inégale sur coûts, variance

Caractéristiques souhaitables d'un système d'échantillonnage

- a. Simple mais flexible, permet des plans différents
- b. Permet de modifier les tailles d'échantillon sans « gaspillage »
- c. Cadre généralisé : éléments et méthodes communs
- d. Utilisation efficace d'une base maintenue à jour
- e. Peut reproduire le plan de sondage de l'EPA
- f. Fait la sélection d'échantillons à 1 et 2 degrés
- g. Donne des échantillons distincts (fardeau de réponse)
- h. Permet de coordonner la collecte des enquêtes
- i. Facilite la transition après remaniement

Stratégie générale – plans à 2 degrés

- Le même ensemble de grappes et de départs sert à toutes les enquêtes
- Les grappes servent d'unités de base à la stratification, l'échantillonnage et la coordination d'échantillons
- Les départs servent à donner des échantillons non chevauchants
- Chaque départ reçoit 2 nombres aléatoires permanents
 - S1 pour la sélection de grappes (\approx PPT, taille = n^{bre} départs)
 - S2 pour la sélection de départs dans les grappes
- Chaque enquête peut stratifier les grappes à sa façon
 - Mais une stratification commune facilite la coordination

Stratégie générale – plans à 2 degrés

- Ex. valeurs S1 des 30 départs des 4 grappes dans une strate :

Grappe 1: 887 192 659 738 130 427 773

Grappe 2: 814 223 518 144 739 051 405 667 944

Grappe 3: 046 409 714 275 151 382 760 482

Grappe 4: 631 309 878 218 919 535

- Les grappes sont choisies selon la séquence des S1 :

046 051 130 144 151 192 218 223 275 309 382 405 409 427 482

518 535 631 659 667 714 738 739 760 773 814 878 887 919 944

- Ou une enquête peut choisir les grappes d'une autre enquête pour coordonner leur collecte
 - Méthode non biaisée si les strates sont les mêmes
- Chaque stratification génère sa propre séquence de S1



Stratégie générale – plans à 2 degrés

- Dans chaque grappe la séquence des valeurs de S2 détermine l'ordre de sélection des départs
- La même séquence de S2 sert à toutes les enquêtes
- Les k (≥ 1) prochains départs sont pris à chaque fois que leur grappe est sélectionnée

- Ex. valeurs S2 des 7 départs dans la grappe 1 :

<u>Départ</u>	:	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>
S2	:	286	506	732	178	564	355	941

donnent l'ordre de sélection de départs 4-1-6-2-5-3-7-4-1-6-2-5-3-7...

- Après l'épuisement de la liste, on resélectionne les départs (normalement après plusieurs années)

Stratégie générale – plans à 2 degrés

Traitement de la croissance (nouveaux logements)

- Les nouvelles unités d'une grappe sont allouées séquentiellement aux départs (comme avec l'EPA)
- On peut aussi générer des nouveaux départs pour maintenir l'homogénéité des tailles des départs :
 - Les nouveaux départs ont des valeurs $S1$ et $S2$
 - Ces valeurs sont insérées dans les séquences en cours pour la strate ($S1$) et la grappe ($S2$)
 - Ces ajouts modifient les probabilités de sélection des grappes, et la séquence de l'échantillon (un ajustement sera nécessaire)



Ex. traitement de la croissance

Répartition de 32 unités *plus 5 « naissances »* aux 7 départs d'une grappe

Départ	1	2	3	4	5	6	7
S2	286	506	732	178	564	355	941
Unité	1	2	3	4	5	6	7
	8	9	10	11	12	13	14
	15	16	17	18	19	20	21
	22	23	24	25	26	27	28
	29	30	31	32	<i>33</i>	<i>34</i>	<i>35</i>
	<i>36</i>	<i>37</i>					

Ordre de départs
initial: *4-1-6-2-5-3-7.*
Révisé : *9-4-1-10-6-2-5-8-3-7.*

Ajout de 16 autres unités (et 3 départs) plus tard à la grappe

Départ	1	2	3	4	5	6	7	<i>8</i>	<i>9</i>	<i>10</i>
S2	286	506	732	178	564	355	941	<i>680</i>	<i>79</i>	<i>351</i>
Unité	1	2	3	4	5	6	7	<i>38</i>	<i>39</i>	<i>40</i>
	8	9	10	11	12	13	14	<i>41</i>	<i>42</i>	<i>43</i>
	15	16	17	18	19	20	21	<i>44</i>	<i>45</i>	<i>46</i>
	22	23	24	25	26	27	28	<i>47</i>	<i>48</i>	<i>49</i>
	29	30	31	32	33	34	35	<i>50</i>	<i>51</i>	<i>52</i>
	36	37	<i>53</i>							

Quelques résultats (voir article)

- n sélections avec S1 donnent \approx ÉPPTAR de n grappes
- Nombre de fois grappe i choisie, $n_i \sim$ Hypergéométrique
- $k=1$, n petit \rightarrow mêmes π_i, π_{ij} qu'un ÉASSR de n départs
- Si $k > 1$ faire comme si les échantillons sont indépendants ?
- Un ajustement est nécessaire si on ajoute N_B nouveaux nombres aléatoires à une séquence (S1 ou S2) de N_X :
 - Prochaine sélection $\in N_B \rightarrow$ retenir avec probabilité π
 - Retenue \rightarrow prendre aussi $n - 1$ prochaines unités
 - Pas retenue \rightarrow seulement prendre n prochaines unités
 - Prochaine sélection $\in N_X \rightarrow$ la garder et $n - 1$ suivantes
 - $\pi = 1 - [(N-2)n] / [(N_X-1)N]$ fera l'affaire (si $\pi \geq 0$)

Premier problème : resélection

- Les départs sont resélectionnés si leur grappe est épuisée (parce que sélectionnée trop souvent et/ou $k > 1$)
- Calcul de probabilités d'épuiser les départs d'une grappe en moins de 5 ans pour diverses fractions de sondage inverses et avec $k = 2$ ou 3 (selon le plan de l'EPA)

FSI = $1/f$	40	60	40	60	80	72	90
N ^{bre} départs	240	360	240	360	480	32	540
Départs/grappe	16	24	12	18	24	18	18
N ^{bre} grappes	15	15	20	20	20	24	30

Prob. qu'une grappe i soit épuisée en moins de 5 ans

Avec $k=2$	0,006	0,000	0,012	0,000	0,000	0,000	0,000
Avec $k=3$	0,182	0,009	0,152	0,018	0,001	0,007	0,002

Deuxième problème : efficacité

- La sélection avec remise peut nuire à l'efficacité du plan
- 6 sélections donnent souvent moins de 6 grappes distinctes
- Avec $\rho \leq 0,24$ l'effet de plan n'est que légèrement affecté
- **Alternative : utiliser des valeurs de S1 équidistantes**

FSI = $1/f$	40	60	40	60	80	72	90
Nbre départs	240	360	240	360	480	432	540
Départs/grappe	16	24	12	18	24	18	18
Nbre grappes	15	15	20	20	20	24	30

Prob. la sélection de $n=6$ grappes en donne n' distinctes

$\Pr(n'=6 n=6)$	0,337	0,330	0,464	0,455	0,450	0,525	0,603
$\Pr(n'=5 n=6)$	0,474	0,474	0,426	0,429	0,431	0,392	0,342
$\Pr(n' < 5 n=6)$	0,189	0,196	0,110	0,116	0,119	0,083	0,055

Coordination d'enquêtes à 2 degrés

- Les enquêtes avec les mêmes strates peuvent sélectionner dans les mêmes grappes sans biais
- Sinon, on peut substituer des grappes échantillonnées au niveau des intersections de strates
- Cas spécial : coordination avec l'EPA
 - L'EPA utilise une base mise à jour 4 fois par an
 - Les autres enquêtes s'y coordonnent 2 fois l'an
 - Les nouveaux départs dérangent la coordination – un nouveau S1 peut avoir un effet domino sur l'échantillon
 - Suggestion : coordonner avec l'échantillon EPA prévu
 - Se servir d'intervalles de S1 pour éviter un effet domino
 - Les nouveaux départs auront un effet nuisible limité

Reproduction du plan de sondage EPA

- On aimerait utiliser le même mécanisme pour reproduire le plan de l'EPA en milieu rural
- Approche :
 - Au lieu de procéder par strate, procéder par groupe de renouvellement dans chaque strate
 - Utiliser des valeurs $S1_{EPA}$ qui se regroupent par grappe
- Au lieu d'avoir (les couleurs représentent les grappes) :

046 051 130 144 151 192 218 223 275 309 382 405 409 427 482
518 535 631 659 667 714 738 739 760 773 814 878 887 919 944

- Faire ceci :

046 051 130 144 151 192 218 223 275 309 382 405 409 427 482
518 535 631 659 667 714 738 739 760 773 814 878 887 919 944

Alternative à l'utilisation de départs

- L'utilisation de départs entraîne du « gaspillage » d'échantillon, et l'épuisement prématuré de grappes
- Au lieu de départs, on peut ordonner les unités de chaque grappe (pseudo-)aléatoirement
- Les valeurs S_2 seront allouées aux unités mêmes – il y en aura autant que d'unités dans la grappe
- Chaque enquête prendra exactement le nombre d'unités voulues dans une grappe
- Le traitement des nouvelles unités (insertion de valeurs S_2 , ajustement nécessaire) a été présenté
- L'estimation de la variance sera approximative ($k > 1$)

Contrôle du chevauchement des enquêtes à 1 et à 2 degrés

- Les enquêtes à 1 degré stratifient au niveau de l'unité
- Microstrate peut servir à contrôler leur chevauchement
- Le chevauchement des enquêtes à 2 degrés est géré
- Pour le chevauchement des enquêtes à 1 et 2 degrés :
 - Une portion fixe, mais mobile, de chaque grappe est inaccessible aux échantillons des enquêtes à 1 degré
 - Cette portion est établie, p. ex., au niveau régional
 - Elle est composée de départs récemment échantillonnés par les enquêtes à 2 degrés et des départs suivants
 - On réserve des étendues de valeurs S2 correspondants à ces départs – ce qui traite également les naissances



Comparaison de trois approches

Caractéristique désirable	Plan de l'EPA	Plan à 1 degré (ÉAS)	Stratégie proposée
a. Simple et flexible	x x x	✓ ✓ x	✓ ✓ ✓
b. Modifier tailles d'éch.	✓ x x	✓ ✓ ✓	✓ ✓ x
c. Système généralisé	✓ x x	✓ ✓ x	✓ ✓ ✓
d. Util. d'une base à jour	✓ x x	✓ ✓ ✓	✓ ✓ x
e. Reproduit plan de l'EPA	✓ ✓ ✓	x x x	✓ ✓ ✓
f. Éch. à 1 et à 2 degrés	x x x	x x x	✓ ✓ ?
g. Échantillons coordonnés	✓ ?	x ?	✓ ?
h. Coordination de collecte	✓ ?	x x x	✓ ?
i. Transition/remaniement	x x x	✓ ✓ ✓	✓ ?

Prise en considération des tendances

- Sérieux impact sur les tendances mensuelles de l'EPA – augmentation de variances jusqu'à 25 %
 - Décision de garder le plan actuel de l'EPA
 - Choix pour les autres enquêtes : se coordonner avec l'EPA ou utiliser la nouvelle approche
- Ajustement à la stratégie :
 - Diviser les départs de chaque grappe en 2 portions fixes
 - Portion EPA : répliquer l'approche EPA
 - Autre portion : appliquer la nouvelle stratégie
- Gestion du chevauchement avec enquêtes à 1 degré
 - Autre portion : approche présentée plus tôt
 - Portion EPA : exclure tous les départs dans la portion EPA ou exclure une portion fixe dans chaque groupe de renouv.

Conclusion

- La stratégie présentée est avant tout un mécanisme de sélection
- Ses atouts principaux :
 - Grande flexibilité
 - Gestion du chevauchement
- Ses lacunes principales :
 - Perte d'efficacité (stratégie de base rejetée pour l'EPA)
 - Besoins d'ajustements (ne peut satisfaire tous les besoins à la fois)



Pour plus d'information,
veuillez contacter :

For more information
please contact:

Jean-Louis Tambay

Jean-Louis.Tambay@statcan.gc.ca