

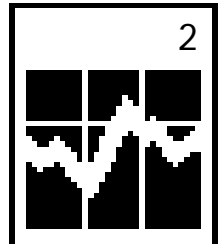
De la méthodologie d'enquête à l'analyse des données première partie

Michel Latouche, Statistique Canada

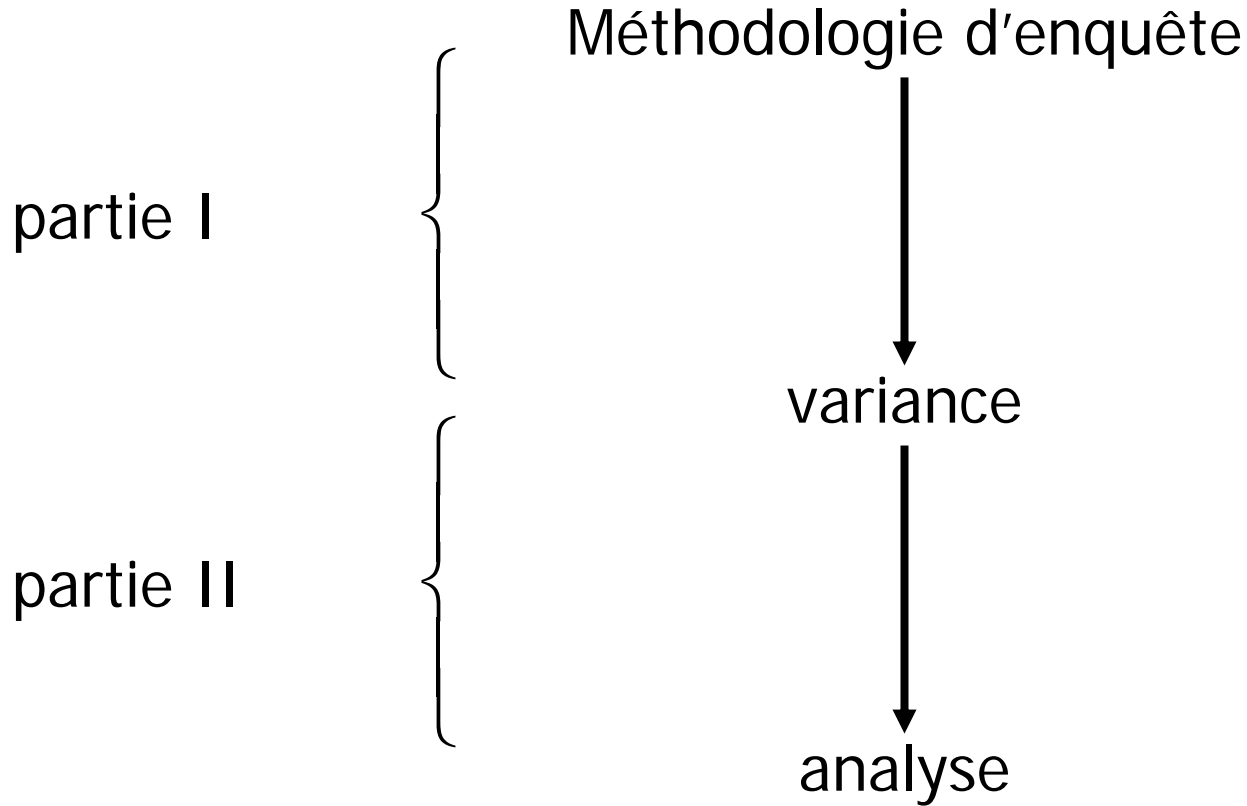
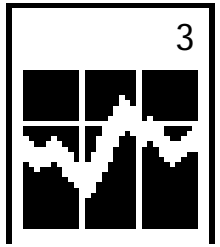
Présentation au CIQSS-INRS-CRSTC
janvier 2002

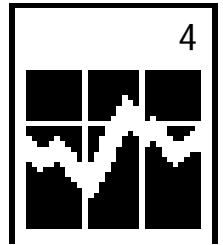
Ont également participé

- Johanne Tremblay
- Sylvie LaRoche
- Serge Godbout
- Gavin Thompson
- Michelle Simard
- Martin Renaud



Préambule



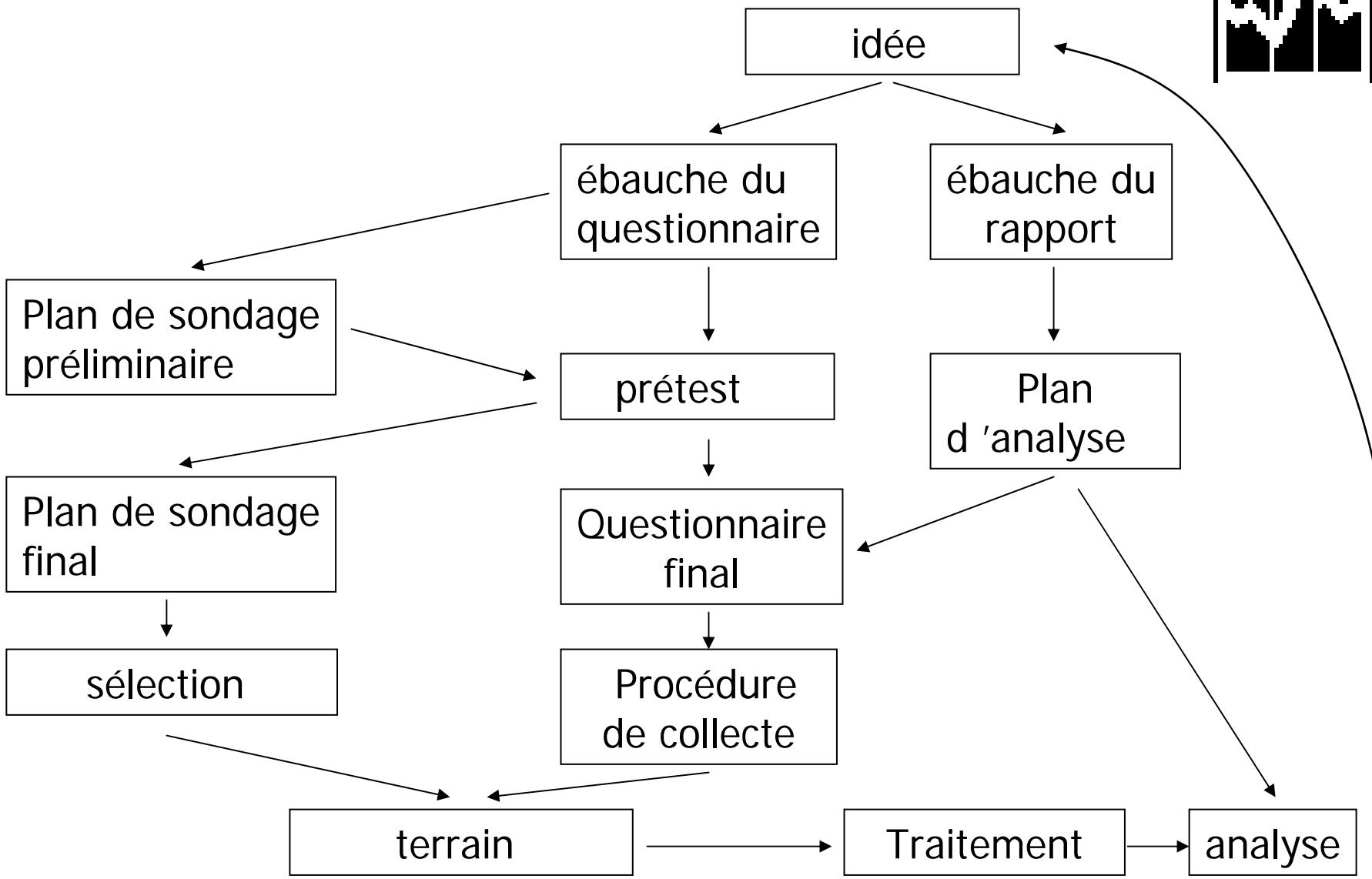
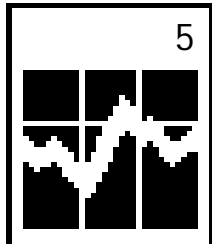


Plan de la session

- Introduction
- Types d'enquête
- Concepts associés à l'échantillonnage
- Plans de sondage et effets de plan
- Estimation de la variance
- Impact du plan de sondage

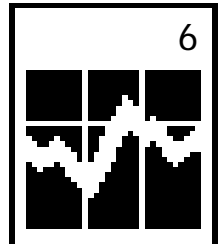
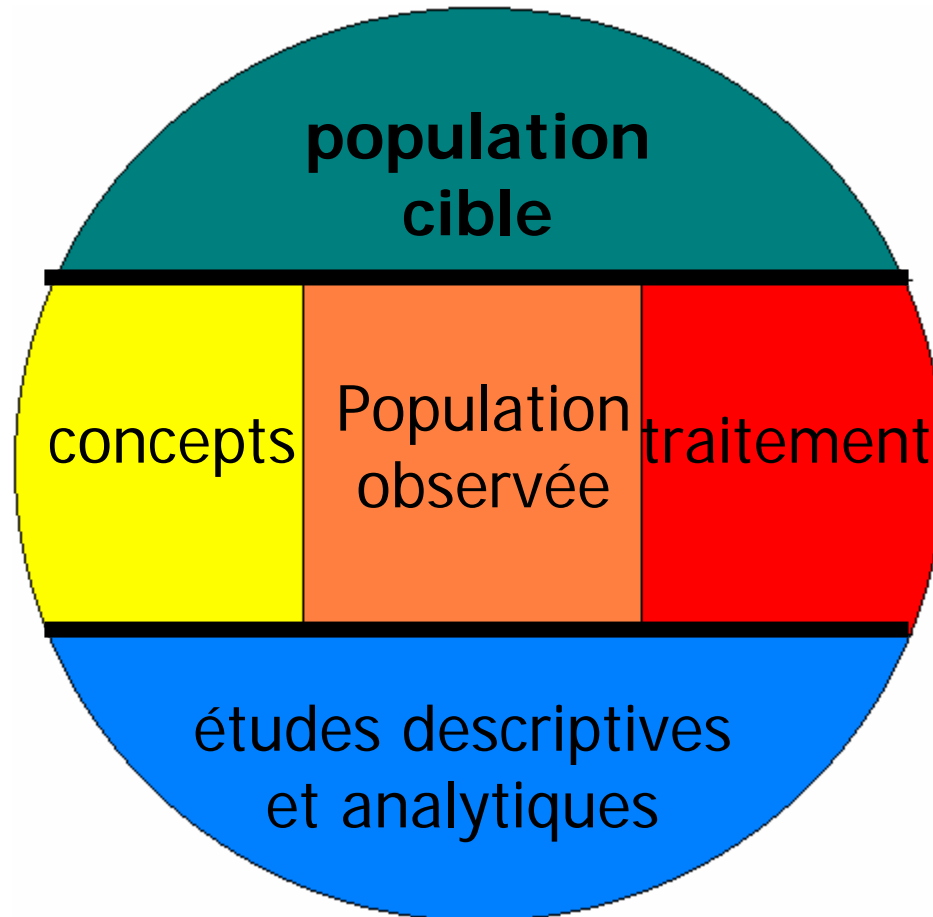
- L'Enquête sur la population active
- La pondération de l'Enquête sur la dynamique du travail et du revenu

Introduction



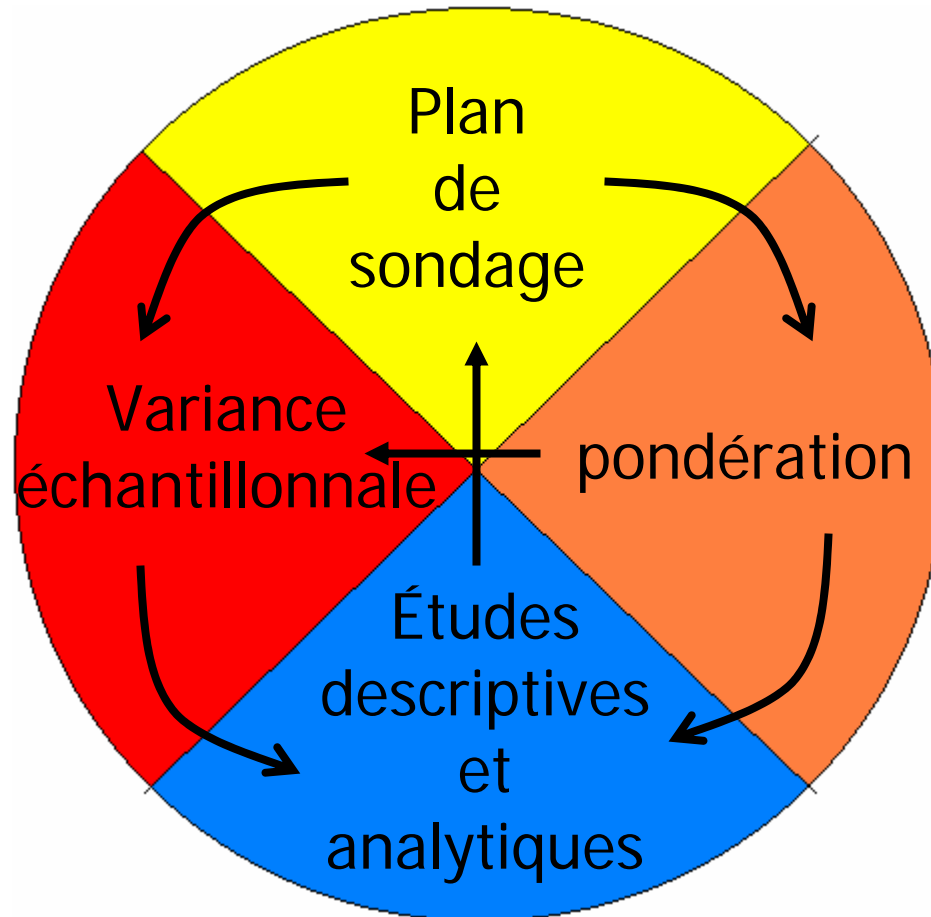
Introduction (suite)

- Méthodologie

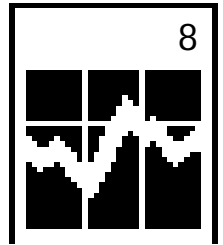


Introduction (suite)

- Échantillonnage



Introduction (suite)



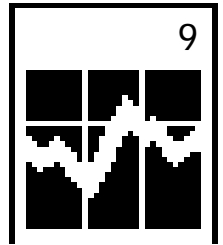
Exemple 1- usage des poids

- Revenu moyen des familles
- Sans poids, on suppose que toutes les unités ont le mêmes poids

famille	revenu (\$)	poids	\$Xpoids
1	15 456	253	3 910 368
2	37 037	159	5 888 883
3	44 281	102	4 516 662
4	12 500	350	4 375 000
5	27 036	292	7 894 512
somme	136 310	1 156	26 585 425
moyenne	27 262		22 998

Introduction (suite)

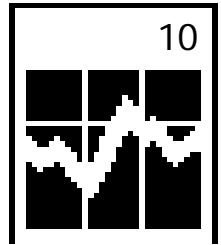
Exemple 2-usage du plan de sondage



observation	poids	revenu (\$)
1	350	29666
2	350	29037
3	350	27207
4	350	27319
5	350	24001
6	350	24763
7	350	27185
8	350	20836
9	350	28596
10	350	28911
11	350	25431
12	350	29359
revenu moyen		26859
écart-type de la moyenne		766
t observée		2,4278
prob(t)		0,0335

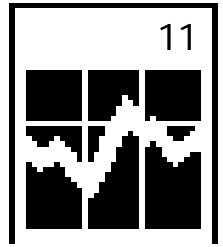
Introduction (suite)

Exemple 2-usage du plan de sondage



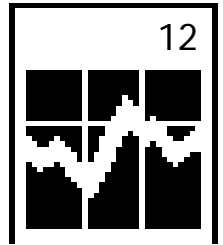
observation	poids	revenu (\$)	grappe	revenu moyen par individu
1	350	29666	1	
2	350	29037	1	
3	350	27207	1	
4	350	27319	1	28307
5	350	24001	2	
6	350	24763	2	
7	350	27185	2	
8	350	20836	2	24196
9	350	28596	3	
10	350	28911	3	
11	350	25431	3	
12	350	29359	3	28074
revenu moyen		26859		26859
écart-type de la moyenne		766		1678
t observée		2,4278		1,1078
prob(t)		0,0335		0,3833

Types d'enquête

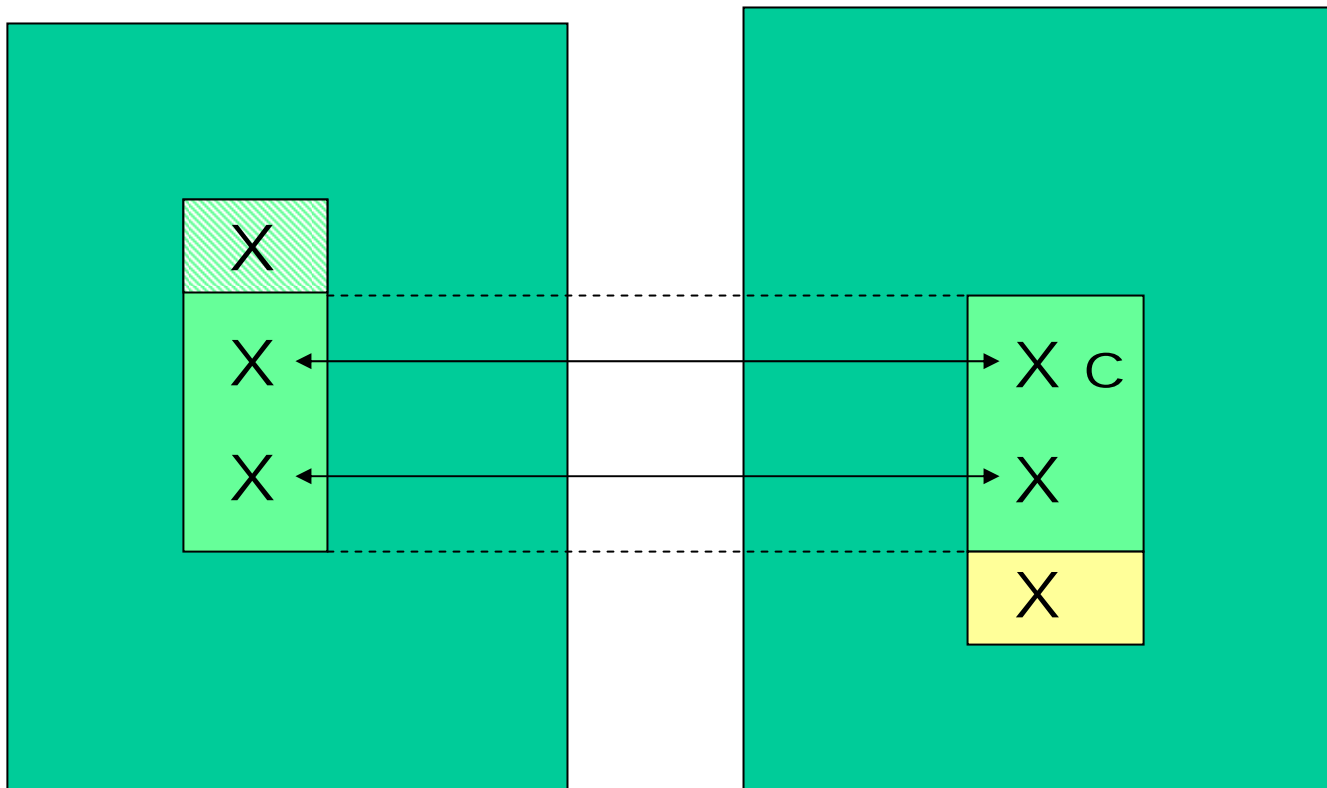


- Occasionnelle
- Répétée indépendante
- Répétée avec chevauchement
- Longitudinale par cohorte
- Longitudinale par panel
- Longitudinale par panel rotatif

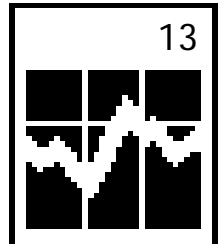
Types d'enquête (suite)



- Enquête longitudinale avec rotation
 - Ex: EDTR et l'EPA



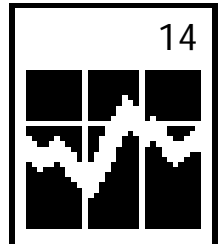
Concepts associés à l'échantillonnage (suite)



- Types de bases
 - Base de liste
 - l'unité statistique correspond à l'unité déclarante
 - l'unité déclarante EST RELIÉE à l'unité statistique
 - Base aréolaire
 - l'unité déclarante fait partie de l'unité statistique

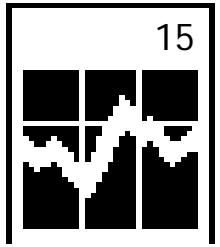
Unité échantillonnale		Unité statistique
entreprise	→	employé
école	→	élève
aire géographique	→	ferme
ménage	→	personne

Concepts associés à l'échantillonnage (suite)



- Variabilité et variance échantillonnale
 - Considérons une population de personnes
 - Études descriptives
 - Mesures de tendance centrale
 - » Moyenne
 - » Médiane

Concepts associés à l'échantillonnage (suite)



- Variabilité et variance échantillonnale (suite)
 - Considérons une population de personnes
 - Mesure de dispersion

» Variance

$$V(y) = \sum_{i=1}^N \frac{(Y_i - \bar{Y})^2}{N}$$

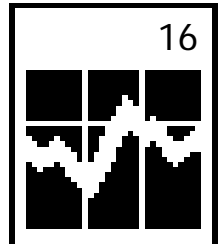
» Écart-type

$$É.T.(y) = \sqrt{V(y)}$$

- Relations

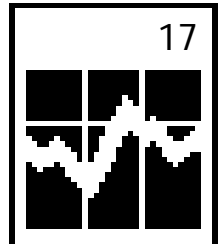
- » Corrélation, régression
- » Fonction de risque
- » Autres

Concepts associés à l'échantillonnage (suite)



- Variabilité et variance échantillonnale (suite)
 - On veut estimer les caractéristiques de la population à partir d'un échantillon
 - Chaque mesure estimée a une marge d'erreur
 - Erreur échantillonnale
 - Variabilité entre les différents échantillons
 - Population fictive composée de tous les échantillons possibles
 - Ces échantillons sont de même taille et sélectionnés de façon similaire et indépendante

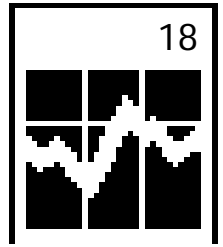
Concepts associés à l'échantillonnage (suite)



- Variabilité et variance échantillonnale (suite)
 - Erreur échantillonnale dépend de
 - Taille de l'échantillon
 - Plan de sondage
 - Variabilité intrinsèque de la caractéristique dans la population
 - » Cette variabilité est estimée par

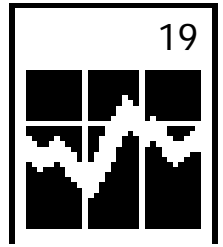
$$s_y^2 = \sum_{i=1}^n \frac{(y_i - \bar{y})^2}{n-1} \quad \text{é.t.}(Y) = s_y$$

Concepts associés à l'échantillonnage (suite)

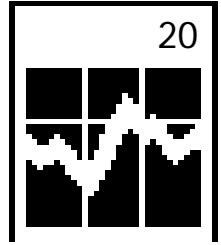


- Variabilité et variance échantillonnale (suite)
 - L'erreur échantillonnale peut être estimée si on connaît les probabilités de sélections des individus et les probabilités jointes (paire d'individus)
 - L'erreur échantillonnale peut être estimée de deux façon:
 - Formule explicite pour les plans de sondage simples
 - Approximation pour les plans plus complexes

Concepts associés à l'échantillonnage (suite)



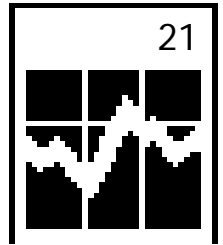
- Pourquoi utiliser des plans complexes
 - Réduction des coûts
 - Absence d'une base de sondage de qualité
 - Facilite la sélection et/ou la collecte
 - Meilleures estimations



Plans de sondage

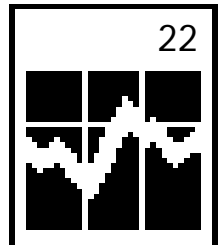
- Simples
 - Bernouilli
 - Poisson
 - Aléatoire simple avec ou sans remise
 - systématique
 - **Stratifié**
 - **Probabilités proportionnelles à la taille**
 - **grappes**
- Complexes
 - **plusieurs degrés**
 - **combinaison de plusieurs plans**

Plan aléatoire simple stratifié



- Une strate est une division de la base de sondage.
- Cas des enquêtes sociales
 - Région géographique:
 - Province, région métropolitaine de recensement, etc.
 - Variables socio-démographiques:
 - Âge, sexe, état matrimonial, etc.
 - Autres variables
 - Niveau de scolarité, type d'emploi, etc.
 - Combinaison de ces catégories

Plan aléatoire simple stratifié (suite)



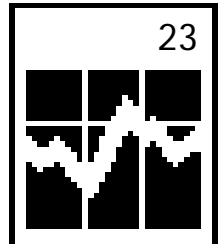
- L'estimation de la variance est similaire au plan simple; on traite chaque strate séparément et on additionne les estimations

$$\hat{Y} = \sum_{h=1}^H \hat{Y}_h$$

où h représente la strate

$$\hat{Y}_h = N_h \bar{y}_h = \sum_{i=1}^{n_h} w_{ih} y_{ih} \quad \text{avec} \quad w_{ih} = N_h / n_h$$

Plan aléatoire simple stratifié (suite)

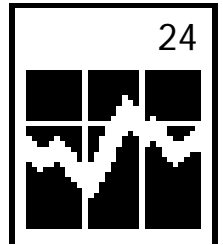


$$\hat{V}(\hat{Y}) = \sum_{h=1}^H \hat{V}(\hat{Y}_h)$$

$$\hat{V}(\hat{Y}_h) = N_h^2 \left(1 - \frac{n_h}{N_h} \right) \frac{s_{y_h}^2}{n_h} \quad \text{sans remise}$$

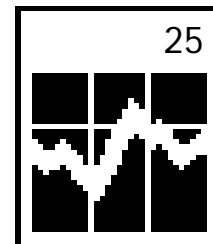
$$\hat{V}(\hat{Y}_h) = N_h^2 \left(1 - \frac{1}{N_h} \right) \frac{s_{y_h}^2}{n_h} \quad \text{avec remise}$$

Plan aléatoire simple stratifié (suite)



- $f_h = \frac{n_h}{N_h}$ est la fraction de sondage
- $w_h = f_h^{-1} = \frac{N_h}{n_h}$ est le poids
- $1-f$ est le facteur de correction pour population finie (finite population correction)
- De façon générale, si f est petit (N très grand) alors les deux formules sont équivalentes

Plan aléatoire simple stratifié (suite)



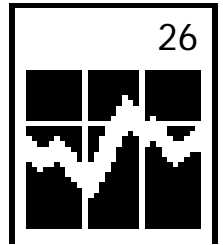
- Effet de plan de sondage :

$$deff = \frac{\hat{V}_{plan\ utilisé}(\hat{Y})}{\hat{V}_{plan\ simple}(\hat{Y})}$$

- Si les strates sont homogènes alors

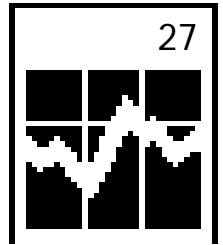
$$\frac{\hat{V}_{plan\ stratifié}(\hat{Y})}{\hat{V}_{plan\ simple}(\hat{Y})} = \frac{\text{variance intra - strate}}{\text{variance totale}} \pi 1$$

Probabilités proportionnelles à la taille



- Les unités n'ont pas toutes la même probabilité d'être choisies
- Les probabilités de sélection sont connues avant la pige
- Probabilités reliées à la taille de l'unité échantillonnale
 - Superficie
 - Taille de l'école
 - Nombre d'employés dans l'entreprise
 - Nombre de personnes dans l'aire géographique

Probabilités proportionnelles à la taille



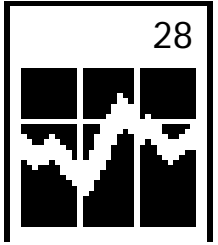
π_i est la probabilité que i soit sélectionnée

π_{ij} est la probabilité que i et j soient sélectionnées

$$\pi_i = kX_i$$

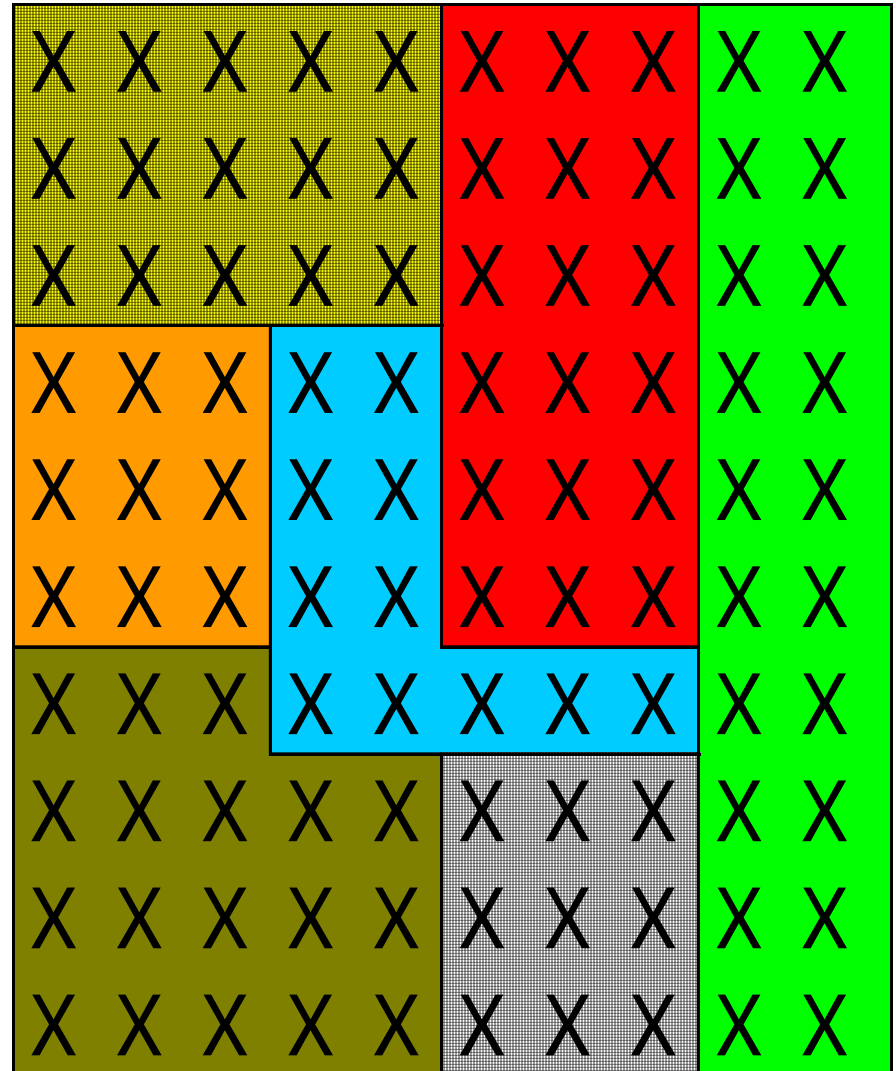
$$\hat{V}(\hat{Y}) = \frac{-1}{2} \sum_{k \neq l} \frac{(\pi_{kl} - \pi_k \pi_l)}{\pi_{kl}} \left(\frac{y_k}{\pi_k} - \frac{y_l}{\pi_l} \right)^2$$

- Contribue à réduire la variance si x est prop. à y

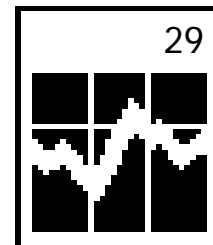


E.A.S. sans remise de grappes

- $N = 7$ grappes
- $n = 2$
- $M_i =$ taille de la grappe i
- Toutes les unités de la grappe sont sélectionnées
- $UPE =$ grappe
- $USE =$ unités dans la grappe



E.A.S. sans remise de grappes (suite)



- Exemples

grappe = commission scolaire

unité interviewée = école

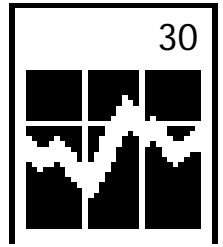
grappe = entreprise

unité interviewée = employé

grappe = pâté de maisons

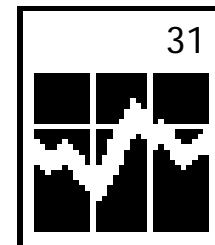
unité interviewée = ménage

E.A.S. sans remise par grappes (suite)



- cas particulier :
 - grappe= regroupement d'unités à interviewer
 - si on choisit une seule grappe, on obtient un plan systématique
- Avantage des grappes
 - Pas besoin de liste d'USE dans les UPE non sélectionnée
 - Réduit coûts de collecte
- Désavantages des grappes
 - Augmente variance (échantillon 'réduit')
 - Complique l'analyse statistique (corrélation entre individus)

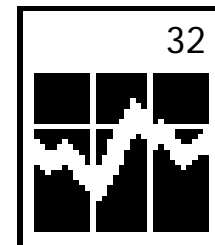
E.A.S. sans remise de grappes (suite)



- Corrélation intra-grappe ou coefficient d'homogénéité
- cas d'un échantillon d'une grappe (systématique)

$$\begin{aligned}\rho &= 1 - \frac{n}{n-1} \frac{SSW}{SST} \\ \delta &= 1 - \frac{NM-1}{NM-a} \frac{SSW}{SST} \\ &= 1 + \frac{NM-1}{NM-a} \frac{n-1}{n} (\rho - 1)\end{aligned}$$

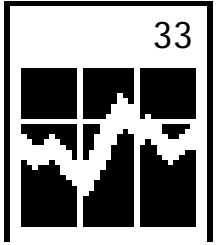
E.A.S. sans remise de grappes (suite)



- Si les grappes sont de tailles M

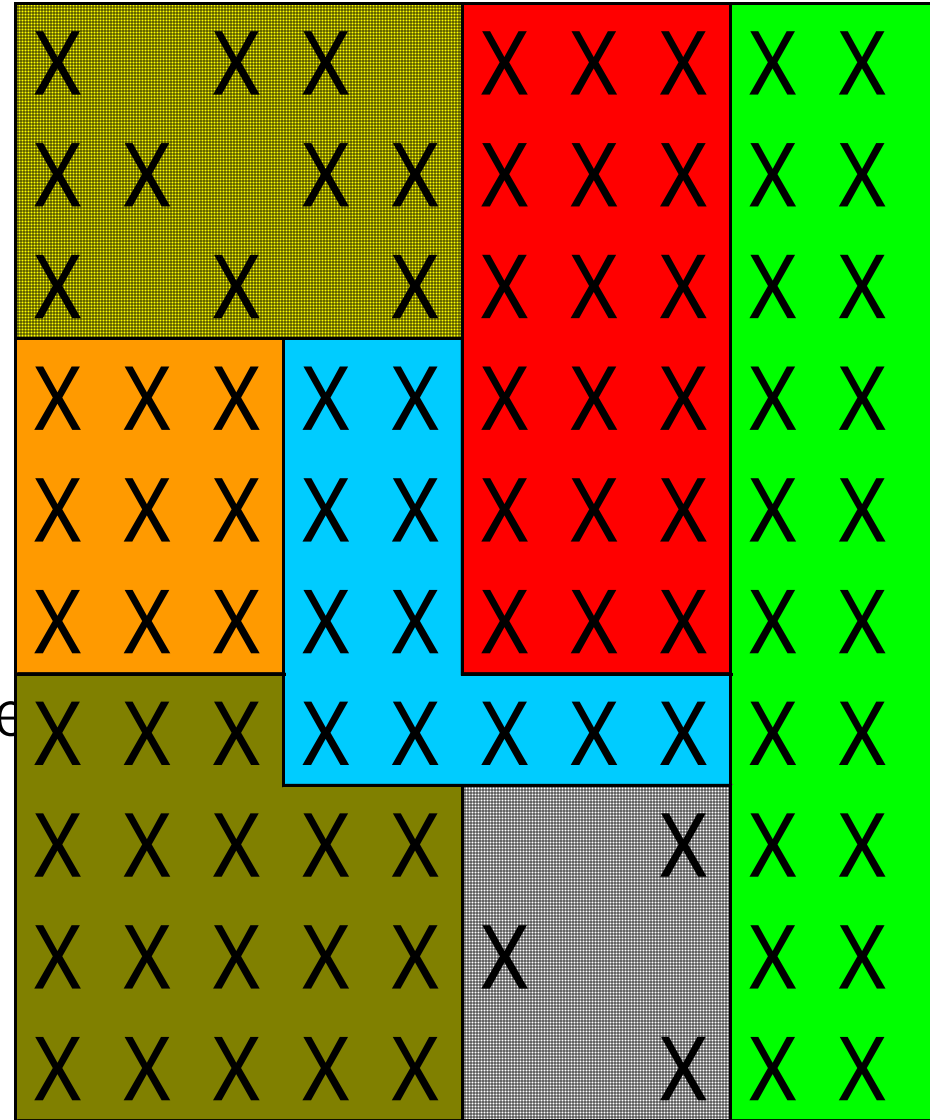
$$deff = \frac{\hat{V}_{é.a.s. \text{ grappe}}(\hat{Y})}{\hat{V}_{plan \ simple}(\hat{Y})} = 1 + (M - 1)\delta \phi$$

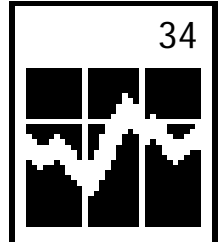
- De façon général $deff$ sera d'autant plus grand si
 - grappes de taille inégale
 - corrélation intra-grappe élevée c.-à-d. les grappes sont homogènes (δ et ρ sont élevés)



Plan à 2 degrés

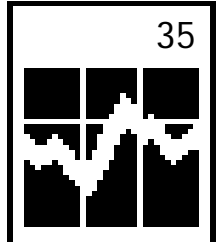
- Unité primaire d'échantillon
UPÉ=grappe
 $n=2$
- Unité secondaire d'échantillon
PSÉ=unité observée
 $m_1=10, M_1=15$
 $m_7=3, M_7=9$
- Variabilité aux deux niveaux de sélection
- Peut être généralisé à 3 degrés ou plus





Plan à deux degrés et plus

- On sélectionne un échantillon de grappe
 - unité primaire d'échantillon (UPE)
- On sélectionne un échantillon à l'intérieur de chaque grappe
 - unité secondaire d'échantillon (USE)
- Le calcul de la variance doit tenir compte de la variabilité échantillonnale à chaque degré.
- Pas toujours possible de dériver explicitement la formule d'estimation de variance
- Calculs toujours très laborieux
- On utilise donc des méthodes approximatives
- $d_{eff} > 1$



Approximation de la variance

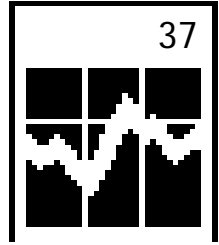
- Méthodes approximatives
 - Développement en série de Taylor
 - Sous-échantillonnage
 - **jackknife**
 - **bootstraps**
 - autres

Approximation de la variance

- Jackknife
 - utilise tout les sous-échantillons de taille $n-1$ obtenus en éliminant tour à tour une UPE.
 - Soit θ la statistique d'intérêt (moyenne, total, ratio, coefficient de corrélation, etc.)

$$\hat{V}_J(\hat{\theta}) = \frac{N-1}{N} \sum_{a=1}^A (\hat{\theta}_{(a)} - \hat{\theta})^2$$

- $\hat{\theta}$ est l'estimation faite à partir de l'échantillon d'origine
- $\hat{\theta}_{(a)}$ est l'estimation faite en ôtant la PSU a



Approximation de la variance

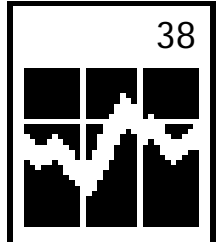
- Bootstraps

- on tire un sous-échantillon de T ($< N$) UPE selon le même design que l'échantillon original. On répète ce processus A fois.

$$\hat{V}_B(\hat{\theta}) = \frac{1}{A-1} \sum_{a=1}^A (\hat{\theta}_a^* - \hat{\theta}^*)^2$$

$$\hat{\theta}^* = \frac{1}{A} \sum_{a=1}^A \hat{\theta}_a^*$$

- $\hat{\theta}_a^*$ est l'estimation faite à partir de l'échantillon a



Approximation de la variance

- Comment cela peut aider les chercheurs?
 - Les enquêtes produisent des poids jackknife ou bootstraps
 - Un chercheur peut avoir accès à ces poids
 - Il exécute son programme d'estimation pour chaque série différente de poids (A séries) et conserve la valeur de son estimation ($\hat{\theta}_{(a)}$ ou $\hat{\theta}_a^*$) et utilise la formule correspondante.

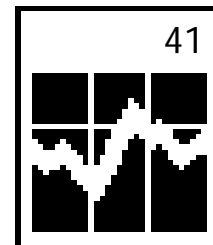
Impact du plan de sondage

- Le plan de sondage affecte les analyses subséquentes de plusieurs façons :
 - contrôle les probabilités de sélection (poids) des unités échantillonnales
 - la variance échantillonnale (d_{eff})
 - la dépendance entre les unités statistiques

Impact du plan de sondage (suite)

- Relation entre le deff et le niveau de l'intervalle de confiance
 - deff élevé implique intervalle trop court

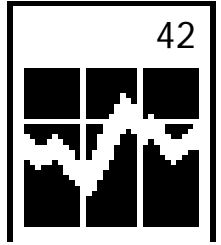
effet de plan (deff)	niveau escompté		
	0,95	0,990	
0,90	0,96	0,993	niveau réel
1,00	0,95	0,990	
1,50	0,89	0,960	
2,00	0,83	0,930	
2,50	0,78	0,900	
3,00	0,74	0,860	



Impact du plan de sondage (suite)

- Relation entre le plan et test d'hypothèse (extrait de SUDAAN)
 - plan à deux degrés stratifié en deux régions
 - UPE sont des écoles (3 par région)
 - $n=310$ étudiants
 - $N \approx 7\,000\,000$
 - variables : sexe et consommation d'alcool durant les 30 derniers jours

consommation	femme	homme	total
oui	1 523 319	1 897 281	3 420 600
non	1 971 831	1 588 849	3 560 680
total	3 495 150	3 486 130	6 981 280



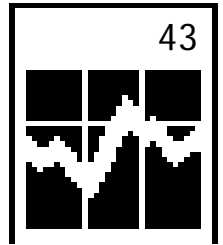
Impact du plan de sondage (suite)

logiciel	% consommateur		Chi-deux	prob.
	femme (é.-t.)	homme (é.-t.)		
SAS	43,58 (0,03)	54,42 (0,03)	82 065,50	0,00
SAS normé	43,58 (3,98)	54,42 (4,00)	3,64	0,06
SUDAAN	43,58 (4,17)	54,42 (5,49)	3,23	0,15

$$w'_i = n \frac{w_i}{\sum_{j=1}^{310} w_j}$$

$$\sum_{j=1}^{310} w'_j = n$$

ENQUÊTE SUR LA POPULATION ACTIVE (EPA)



- OBJECTIF PRINCIPAL
 - Produire des estimations sur l'emploi et le chômage à l'échelle nationale et régionale
- VENTILATION
 - régions infra-provinciales (RÉAE), (RMR)
 - branche d'activité, profession
 - durée (de l'emploi ou du chômage)
 - nombre d'heures travaillées,...

EPA (suite)

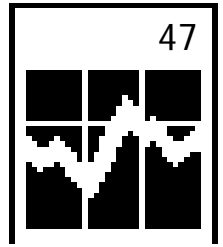
- POPULATION CIBLE
 - Population canadienne à l'exclusion des personnes
 - vivant dans les réserves indiennes, sur les terres de la couronne
 - des Territoires du Nord-Ouest, du Nunavut
 - Pensionnaires d'institutions
 - Membres à temps plein des forces armées
 - 98 % de la population canadienne
- Statut d'emploi de tous les membres du ménage de 15 ans et +

EPA (suite)

- PLAN D'ÉCHANTILLONNAGE
 - Enquête par panel avec renouvellement de 1/6 de l'échantillon par mois
 - Échantillon à deux degrés choisi à partir d'une base aréolaire (pour la majorité des régions)
 - Au premier degré, on choisit des aires géographiques
 - Au deuxième degré, on choisit un échantillon de logements
 - Les ménages éligibles (composés de personnes de la population cible) des logements choisis constituent l'échantillon
- TAILLE D'ÉCHANTILLON:
 - environ 54 000 ménages
 - environ 130 000 personnes

EPA (suite)

- GÉOGRAPHIE DE L'EPA:
 - Division de recensement (DR) : régions géographiques intermédiaires entre les municipalités et la province établi par les lois provinciales (288)
 - Subdivision de recensement (SDR) : correspond généralement aux municipalités (5984)
 - Secteurs de dénombremments (SD) : unités géographiques correspondant à la charge de dénombrement d'un recenseur (49361)
 - Régions économiques (RE): groupe de DR servant à l'analyse de l'activité économique régionale (72)



EPA (suite)

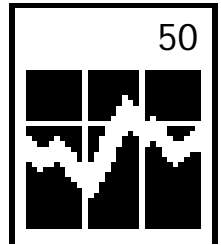
- Régions économiques d'assurance-emploi (RÉAE):
Découpage pour l'assurance-emploi utilisé par DRHC
(55)
- Régions métropolitaines de recensement (RMR) : grands centres urbains dont le noyau compte au moins 100 000 habitants (25)
- Régions urbaines : région avec une population minimale de 1000 habitants et une densité d'au moins 400/km carré (929)

EPA (suite)

- BASE LISTE D'APPARTEMENTS:
 - Liste d'immeubles d'appartements créée à partir d'information provenant de la SCHL
 - Définition d' immeuble d'appartements: immeuble comptant 5 étages habités et au moins 30 logements
 - Permet une meilleure représentativité des locataires d'appartements et minimise l'effet de croissance importante des grappes dans une base aréolaire lorsque de nouveaux immeubles à appartements sont construits dans la grappe
 - Certaines villes seulement où il y a un nombre suffisant d'immeubles pour former une base séparée (18 villes)

EPA (suite)

- STRATIFICATION SELON LES UNITÉS GÉOGRAPHIQUES STANDARD
 - Régions économiques (RE) - 72 RE
 - Régions économiques d'assurance emploi (RÉAE) - 55 RÉAE
 - Stratification par RE X RÉAE - 133 strates -
 - Stratification des RE X RÉAE par secteur urbain, ruraux, régions éloignés



EPA (suite)

- STRATIFICATION OPTIMALE DE LA BASE ARÉOLAIRE
 - Formation de strates plus détaillées sans égard aux contraintes géographiques appliquées à l'intérieur de strates géographiques plus grandes
- Répartit les unités en strates aussi homogènes que possibles selon certaines variables tels que:
 - revenu total
 - occupation
 - niveau d'éducation
 - loyer brut total
 - nombre de logements en propriété
 - groupe d'âge
 - ménage par taille
 - langue maternelle
- 1232 strates au total

EPA (suite)

- STRATIFICATION DE LA BASE D'APPARTEMENTS
- Dans les 18 grandes villes avec base d'appartements,
 - Lorsque le nombre le permet, on la divise en deux:
 - Immeubles d'appartements à faible revenu (revenu moyen des ménages < 20 000\$)
 - Immeubles réguliers
 - On subdivise géographiquement (si possible)
 - On subdivise selon la taille de l'immeuble (si possible)
 - Nombre total de strates varie de 1 à 16 selon les villes

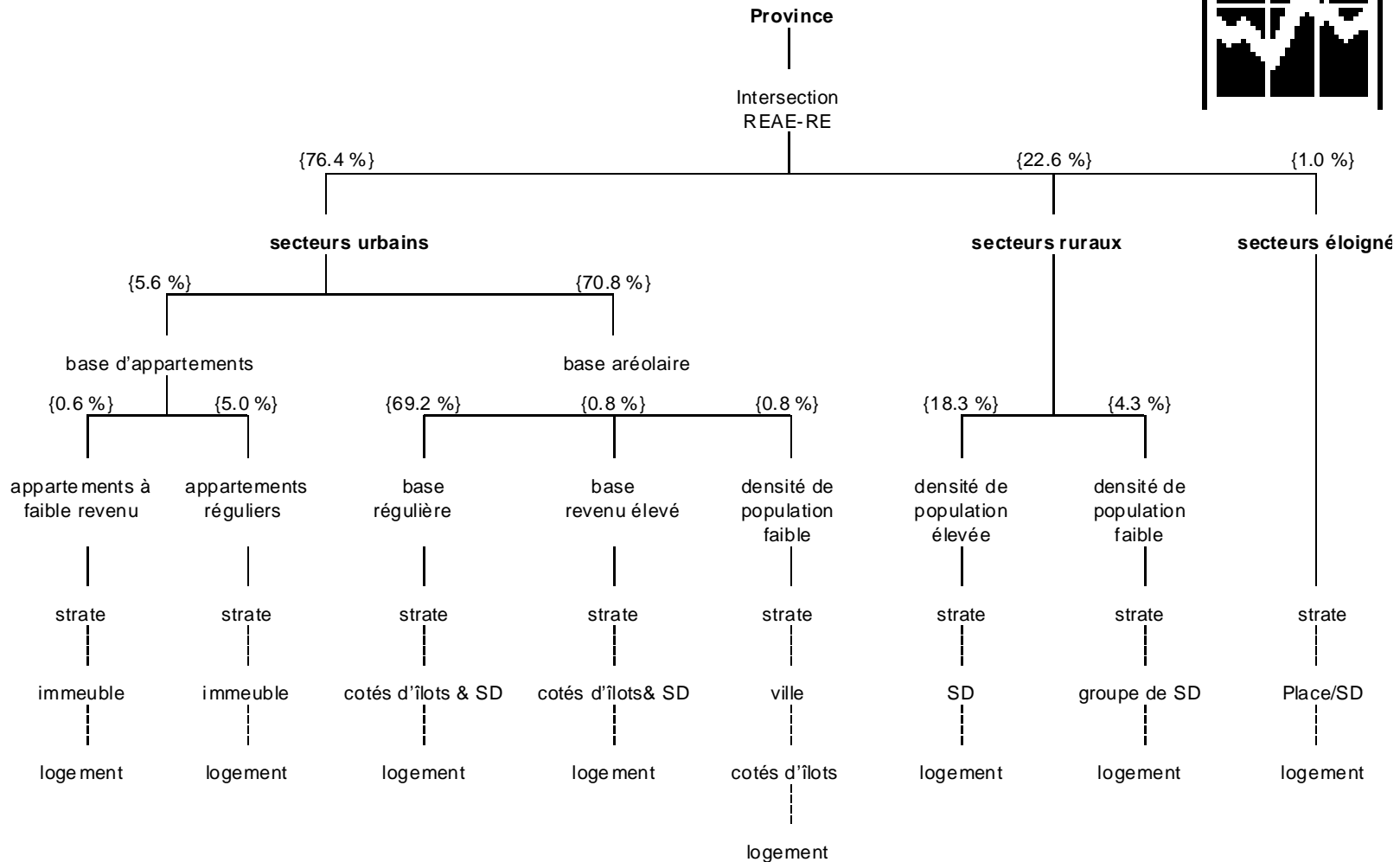
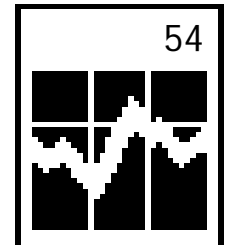
EPA (suite)

- STRATES À REVENU ÉLEVÉ DE LA BASE ARÉOLAIRE
 - Créées dans certaines grandes villes seulement (9)
 - SD avec un revenu moyen d'environ 100 000\$ ou plus
 - 1 à 3 strates selon la ville
 - Vise à diminuer le problème de la représentativité des ménages à revenu élevé
 - Permet de comparer les taux de réponse pour les ménages à revenu élevé
 - Comme c'est une stratification d'aires géographiques dans certaines villes seulement, les ménages à revenu élevé ne se trouvent pas nécessairement dans les strates à revenu élevé

EPA (suite)

- NOMBRE TOTAL DE STRATES
 - 954 strates urbaines dans la base aréolaire
 - 210 strates rurales dans la base aréolaire
 - 68 strates dans la base appartement
 - TOTAL : 1232 strates
 - Strate finale : plus bas niveau, plus petite strate

Plan de sondage de l'Enquête sur la population active - 1995+



| = niveau de stratification
 REAE - Rég. économique d'assurance-emploi
 RE - Région économique
 SD - Secteur de dénombrement
 { } = degré d'échantillonnage
 { % } = pourcentage de l'échantillon total

EPA (suite)

- RÉPARTITION DE L'ÉCHANTILLON
- Échantillon de base de 44 000 ménages
 - On cherche principalement à avoir de bonnes estimations provinciales
 - pour améliorer les estimations au niveau des RÉAE
 - Répartition des RE aux RE x RÉAE proportionnelle à la taille
- Échantillon payé par la DRHC de 10 000 ménages pour améliorer les estimations au niveau des RÉAE
 - Répartition pour maximiser l'amélioration des CV du chômage dans les RÉAE- minimum 600 ménages par RÉAE

EPA (suite)

- Premier degré d'échantillonnage
 - La formation des UPE varie selon les secteurs. Les principaux types de UPE sont:
 - Secteurs urbains de grande taille (50 000 et +)
 - Les autres secteurs urbains
 - SD, partie d'un SD, groupement de SD
 - taille varie entre 150 et 300 logements

EPA (suite)

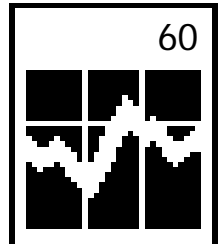
- Sélection des UPE
 - La méthode de sélection varie selon les secteurs.
 - On choisit généralement 6 UPE par strate finale.
 - Les 3 principales méthodes pour sélectionner les UPE sont:
 - systématique PPT avec classement aléatoire dans 6 groupe
 - systématique PPT avec classement aléatoire
 - systématique PPT

EPA (suite)

- Sélection des USE
 - Pour toutes les grappes sélectionnées, on établit la liste de logements
 - On choisit un échantillon systématique de logements à partir de la liste
 - On génère un nombre aléatoire entre 1 et la fraction de sondage inverse de la grappe (FSI)

EPA (suite)

- Taille d'échantillon dans les grappes varie selon le type de strates (nombre espéré de logements)
 - 6 à Toronto, Montréal et Vancouver
 - 8 dans les autres secteurs urbains de grande taille
 - 4 dans les strates à revenu élevé
 - 5 dans les strates de la base d'appartements
 - 3 à 10 dans les strates des autres secteurs urbains
 - 10 dans les secteurs ruraux à plus grande densité
 - 60 dans les secteurs ruraux à plus faible densité
 - 12 dans les régions éloignées



EPA (suite)

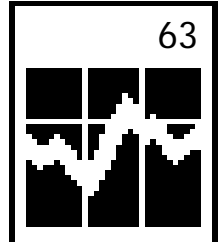
- Rotation
 - 1/6 de l'échantillon est renouvelé à chaque mois (1 groupe de rotation sur les 6)
 - On procède au renouvellement des ménages de la grappe c'est-à-dire qu'un nouvel échantillon de logements généralement de la même grappe viendra remplacer l'ancien et de nouveaux ménages seront interviewés
 - Les grappes demeurent dans l'échantillon plus longtemps
 - Pour environ 90% des renouvellements, le nouvel échantillon provient de la même grappe que celui remplacé

EPA (suite)

- Rotation
 - On procède au renouvellement des ménages de la grappe c'est-à-dire qu'un nouvel échantillon de logements généralement de la même grappe viendra remplacer l'ancien et de nouveaux ménages seront interviewés
 - Les grappes demeurent dans l'échantillon plus longtemps
 - Pour environ 90% des renouvellements, le nouvel échantillon provient de la même grappe que celui remplacé

EPA (suite)

- Poids de base
 - Poids de base = R_h pour tout ménage de la strate h
 - PLAN AUTOPONDÉRÉ DANS CHAQUE STRATE
 - facilite la répartition du travail
 - ajustement de la taille de l'échantillon
 - POIDS DE SONDAGE =
Poids de base x Poids de stabilisation x Sous-poids de grappe



EPA (suite)

- Ajustement de nonréponse
- Dans l'EPA, on utilise
 - RÉAE
 - Type de secteur (ex. base appartement, SD urbain, plan 3 degrés,.....)
 - Groupe de rotation
 - Strates à revenu élevé
- SOUS-POIDS =
Poids de sondage x ajustement pour la non-réponse =
 $w / \text{taux réponse}_b$

EPA (suite)

- Calage
 - Ajuster les poids de sorte à ce que les estimations de l'enquête correspondent à des totaux connus provenant de source externe (contrôles)
 - Permet la concordance avec les sources externes et réduit les différences entre différentes sources d'estimations
 - Si les variables de calage sont corrélées avec les caractéristiques à estimer à réduction de la variance
 - Permet un réajustement de l'erreur de couverture
 - (taux de glissement variant généralement entre 8 et 9% pour le nombre de personnes de 15 ans et +)

EPA (suite)

- Calage
- le calage consiste à la stratification a posteriori simultanée selon plusieurs variables
- C'est une estimation par régression où les variables auxiliaires sont des variables indicatrices d'appartenance à une strate a posteriori
- Variante de l'ERG qui garantit que toutes les personnes du même ménage auront le même poids
Voir Lemaître et Dufour (1987)

EPA (suite)

- Variables de calage
 - Sources: Estimations post-censitaires de population
 - 30 groupes d'âge X sexe par province, RÉ et RMR
 - Nombre de personnes faisant partie de ménages de taille 1, 2 et 3+ par province
 - Nombre de personnes faisant partie de familles économiques de taille 1, 2 et 3+ par province

EPA (suite)

- POIDS FINAL = sous-poids * facteur g
Ou encore,
- POIDS FINAL = poids de base x poids de stabilisation x sous-poids d'échantillonnage x facteur de compensation de non-réponse x facteur g
- DÉFINITIVEMENT PAS UN PLAN SIMPLE

EPA (suite)

- Beaucoup d'enquêtes utilisent la base de sondage de l'EPA
 - Enquête sur la sécurité financière
 - Enquête sur les dépenses des ménages
- D'autres sont mêmes des sous-échantillons de l'EPA ou des enquêtes supplémentaires
 - Enquête sur la dynamique du travail et du revenu
 - Enquête longitudinale nationale sur les enfants et les jeunes
 - Enquête sur les jeunes en transition
 - Enquête nationale sur la santé de la population