L’utilisation des données transversales des enquêtes sur la santé de Statistique Canada (ESCC & ENSP)

François Brisebois, Patrice Mathieu, Mario Bédard
Statistique Canada

Plan de l'atelier

- Survol rapide des enquêtes
- Contexte général d'une analyse statistique
- Comment traiter les valeurs manquantes
- Survol de certains types d'analyse transversale
- Comparaison de populations
  - Standardisation âge-sexe
  - Mises en garde concernant la comparaison d'estimations transversales dans le temps
- Survol de logiciels pour l'analyse de données
- Références générales pour les aspects couverts

Module 1

Survol rapide des enquêtes

Outline

- Le programme des enquêtes sur la santé de STC
- Terminologie
- Enquête nationale sur la santé de la population (ENSP)
- Enquête sur la santé dans les collectivités canadiennes (ESCC)

Programme des enquêtes sur la santé

1994-1999

Enquêtes sur la santé

ENSP ménage	ENSP Nord	ENSP Établissements
Longitudinale & Transversale (Provinces)	Longitudinale & Transversale (Territoires)	Longitudinale & Transversale (-C3) (Étab. santé provinciaux)

Depuis 2000

Enquêtes sur la santé

ENSP ménage	ENSP Nord	ENSP Établissements	ESCC
Longitudinale (Provinces)	Longitudinale (Territoires)	Longitudinale (Étab. santé prov.)	Transversale (Provinces + Territoires)

Terminologie

Longitudinal versus transversal

- **Transversal:**
  - **Enquête une population d'interêt à un temps donné**

- **Longitudinal:**
  - **Enquête une population d'intérêt de façon répétée dans le temps**

Terminologie

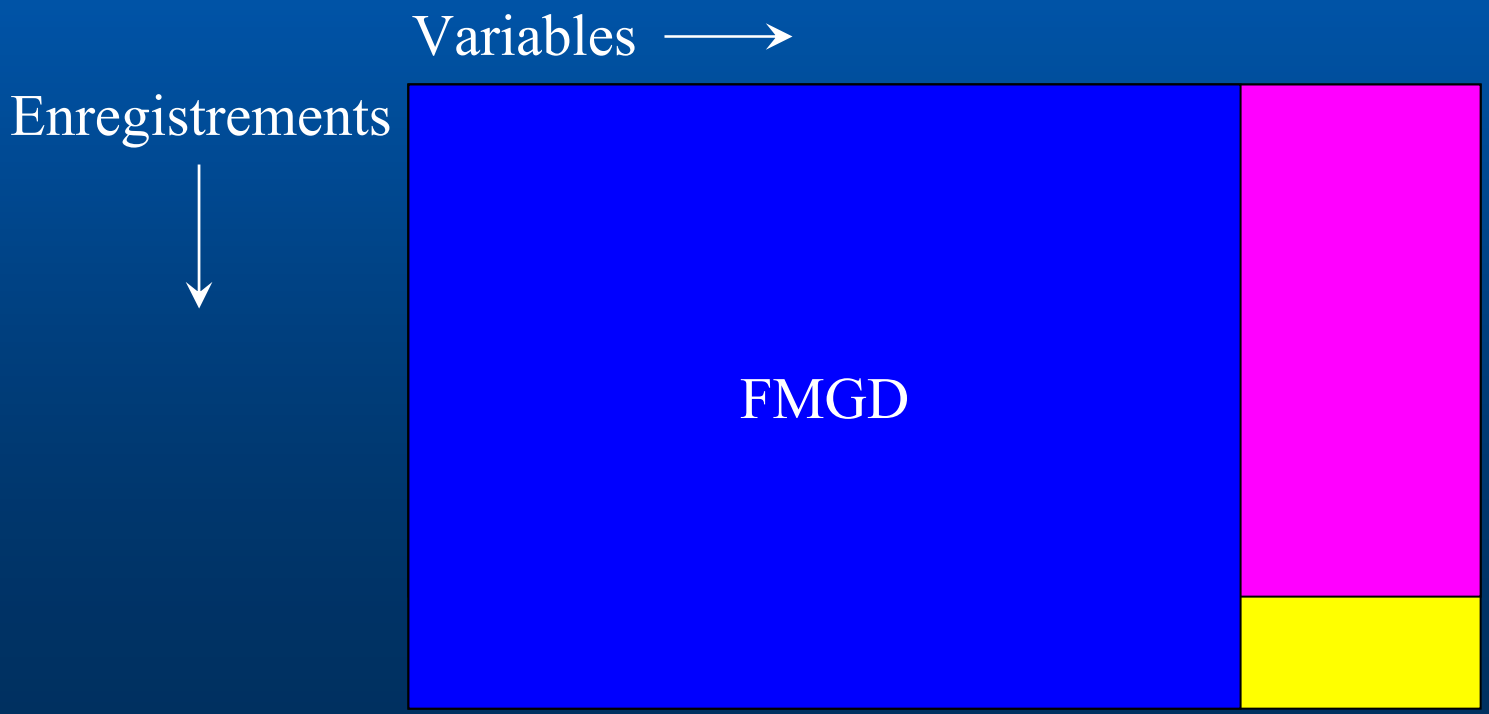
Fichiers maître, partagé, FMGD, fictif

- **Maître:** Fichier contenant toutes les variables pour tous les répondants
- **Partagé:** Fichier contenant toutes les variables mais seulement pour les répondants ayant accepté le partage *(sous-enemble d'enregistrements)*
- **FMGD:** Fichier contenant un sous-ensemble de variables pour tous les répondants *(sous-ensemble de variables)*
- **Dummy:** Version synthétique du fichier maître (à des fins de testing / télé-accès)

Terminologie

Fichiers maître, partagé, FMGD, fictif

- Illustration Maître vs. Partagé vs. FMGD

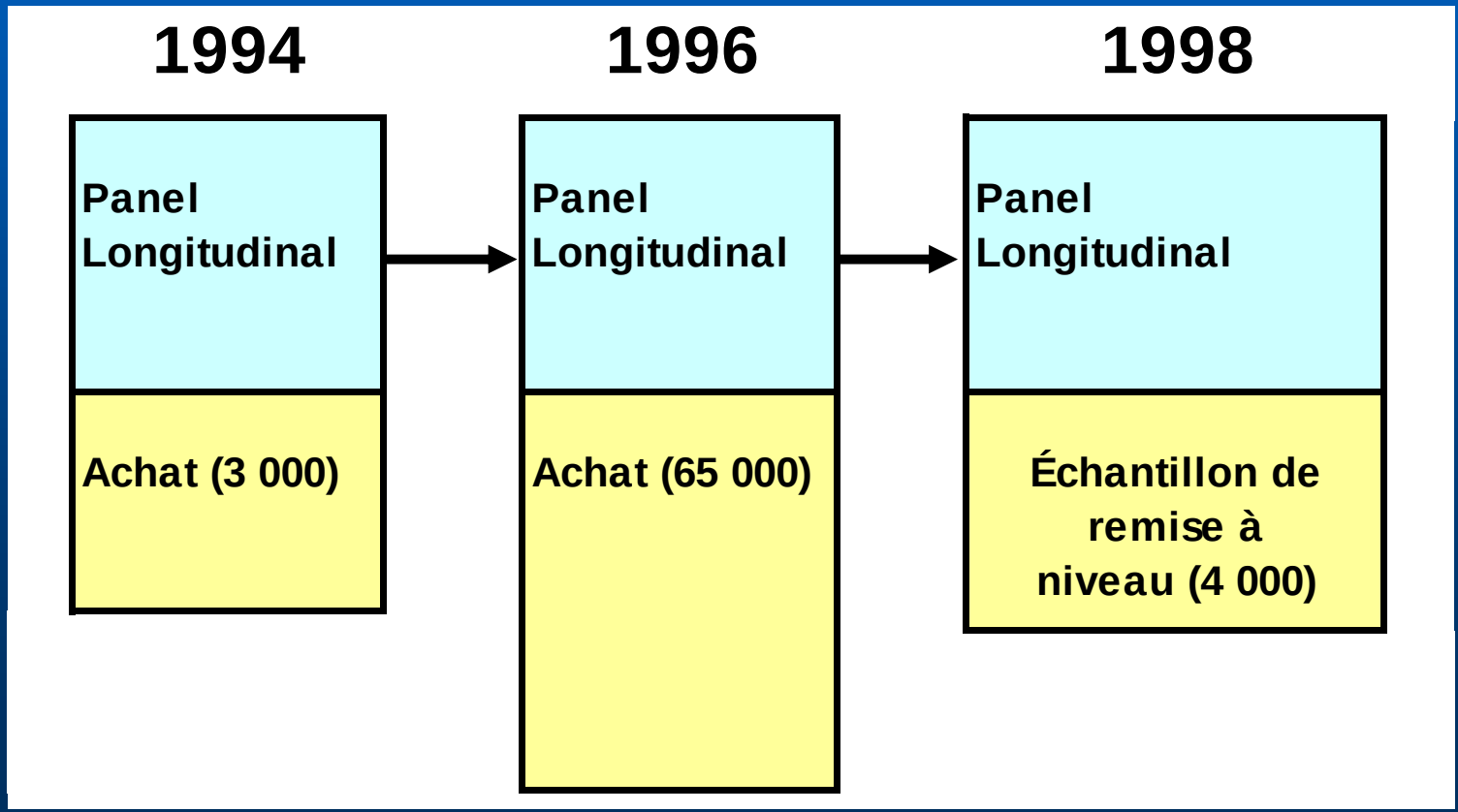


Enquête nationale sur la santé de la population (ENSP)

- **Enquête longitudinale (Mais a aussi servi à des fins transversales pour les 3 premiers cycles)**
- **Panel = 17,276 répondants en 1994**
- **Collecte a débuté en 1994-95 (Cycle 1); le panel est recontacté à tous les deux ans pendant 20 ans**

Enquête nationale sur la santé de la population (ENSP)

- Composition des échantillons transversaux



Enquête nationale sur la santé de la population (ENSP)

- **Situation présente:**
  - **Cycle 4 (2000-01):**
    - Fichier longitudinal a été diffusé en mai 2002
  - **Cycle 5 (2002-03):**
    - Collecte cmplétée; traitement des données
    - Contenu focus: Sommeil, nutrition, histoire résidentielle, marque de cigarette
    - Diffusion du fichier prévue pour le printemps 2004
  - **Cycle 6 (2004-05):**
    - Collecte débutera en juin 2004

Enquête sur la santé dans les collectivités canadiennes (ESCC)

- Enquête transversale
- Cycle de deux ans:
  - 1ère année (.1 – composante régionale):
    - Région de santé – Échantillon 130K
    - Contenu général (avec modules optionnelles)
  - 2ème année (.2 – composante provinciale)
    - Province – Échantillon 30K
    - Contenu spécifique

Enquête sur la santé dans les collectivités canadiennes (ESCC)

- **Situation présente:**
  - **Cycle 1.1 (2000-01):**
    - Fichiers mâitre & partagé diffusés en mai 2002
    - Fichier de microdonnées à grande diffusion (FMGD) diffusé en janvier 2003
  - **Cycle 1.2 (2002):**
    - Santé mentale
    - Achats d'échantillon pour l'Ontario & la Nouvelle-Écosse - > Échantillons représentatifs à l'échelle des régions sous-provinciales
    - Fichiers maître & partagé diffusés en sept. 2003
    - FMGD -> Hiver 2004

Enquête sur la santé dans les collectivités canadiennes (ESCC)

- **Situation présente (suite):**
  - **Cycle 2.1 (2003):**
    - **Collecte jusqu'à la fin de 2003**
    - **Changements dans les frontières géographiques**
    - **Achats d'échantillon pour 3 régions du Québec (représentativité à l'échelle des CLSC)**
  - **Cycle 2.2 (2004):**
    - **Nutrition**

Enquêtes sur la santé (ESCC + ENSP)

- **Sommaire des fichiers transversaux pour la population des ménages**

Année	Enquête	Âge	Maître & FMGD	Partagé
1994-95	ENSP – C1	0+	Général / Santé	Général / Santé
1996-97	ENSP – C2	2+ *	Général / Santé	Général / Santé
1998-99	ENSP – C3	0+	Général / Santé	Général / Santé
2000-01	ESCC 1.1	12+	Santé	Santé
2002	ESCC 1.2	15+	Santé	Santé

Matériel utilisé pour l'atelier

Matériel

- Données fictives
  - Fichier fictif de l'ESCC, Cycle 1.1
- Information inclus
  - Données (ASCII) avec les clichés d'enregistrements
  - Dictionnaire de données
  - Documentation
  - Bootstrap

Lecture les données
SPSS / SAS

- Fichier ASCII -> Cliché d'enregistrement
- Grand nombre de variables
- Utiliser les fichiers périphériques SPSS / SAS fournis
  - Cliché d'enregistrement
  - Étiquettes des variables
  - Étiquettes des valeurs

Lecture les données
SPSS / SAS

- **Répertoires informatiques utilisés:**
  - **Répertoire personnel sur le réseau**
    - **\\ciqss-s2\utilisateurs$\formation\ordinateur xx**
      - **\Data**
      - **\Layout**
      - **\Bootstrp\Data**
      - **\Bootstrp\Layout**

Lecture les données
SPSS / SAS

- **Lire les données dans SPSS / SAS**
  - **Utiliser le programme-gabarit fourni pour lire l'ensemble des enregistrements**
  - **Ensuite, conserver seulement les enregistrements de la région d'intérêt**

Module 2

Contexte général d'une analyse statistique

Outline

- Ai-je assez d’échantillon?
- Estimation d’une statistique
- Estimation de la précision de la statistique

Ai-je assez d'échantillon?

- **Désire que l'échantillon analysé soit représentatif**
  - **Que veux-t-on dire par représentatif?**
    - **Produire des estimations sans biais, et montrant une précision acceptable**
      - **Sans biais: Couverture adéquate de la population étudiée**
      - **Précision: est fonction (entre autres) de la taille de l'échantillon et de la magnitude de la proportion examinée**

Ai-je assez d'échantillon?

- L'ENSP a été planifiée pour garantir une bonne représentativité pour 10 groupes d'âge-sexe, à l'intérieur de chaque province
  - Groupes d'âge: 0-11, 12-24, 25-44, 45-64, 65+
  - Représentativité régionale pour les provinces ayant acheté de l'échantillon additionnel aux cycles 1 et 2

Ai-je assez d'échantillon?

- **Pour l'ESCC,**
  - **Cycle 1.1:**
    - **À l'intérieur de chacune des 136 régions socio-sanitaires**
    - **10 groupes d'âge-sexe (12-19, 20-29, 30-44, 45-64, 65+)**
  - **Cycle 1.2:**
    - **Provincial; régional pour l'Ontario et la Nouvelle-Écosse**
    - **8 groupes d'âge-sexe (15-24, 25-44, 45-64, 65+)**

Ai-je assez d'échantillon?

- **For CCHS,**
  - **Cycle 2.1:**
    - **À l'intérieur de chacune des 133 régions socio-sanitaires (RSS)**
    - **(Possiblement) à l'intérieur de chaque CLSC pour 3 RSS du Québec**
    - **10 groupes d'âge-sexe (12-19, 20-29, 30-44, 45-64, 65+)**
  - **Cycle 2.2:**
    - **À l'intérieur de chaque province (à l'intérieur des régions sous-provinciales si achat d'échantillon)**
    - **15 groupes d'âge-sexe**
      (<1*, 1-3*, 4-8*, 9-13, 14-18, 19-30, 31-50, 51-70, 71+)

Ai-je assez d'échantillon?

- **Que puis-je vérifier?**
  - **Est-ce que la population-cible de l'enquête inclus la population étudiée?**
    - **Référer à la documentation pour connaître la définition de la population-cible**
  - **Est-ce que la précision est bonne ?**
    - **Habituellement, on ne sait pas à l'avance**
    - **Scénarios:**

      Taille d'échantillon (grande ou petite) VS.
      Proportion étudiée (grande ou petite)
    - **Les tableaux de Tables peuvent être utiles**

Ai-je assez d'échantillon?

- **Scénarios (en référence à l'ESCC 1.1) :**
  - **Taux de tabagisme pour une RSS**
  - **Taux de tabsgisme pour les adolescents au Canada**
  - **Alzheimer au Canada**
  - **Alzheimer pour une RSS**
  - **Diabète (CCCA_101 = 1) pour une RSS (GEOA_HR4)**
    - **Obtenir le compte et la proportion dans l'échantillon**

Ai-je assez d'échantillon?

- **Pour les analyses à partir du FMGD, les lignes directrices exigent une taille d'échantillon minimale de 30**
  (voir la section 10.4 du guide d'utilisation du FMGD)
- **Pour les fichiers maître et partagé, la précision de l'estimation est le critère (doit être calculée – avec le bootstrap)**
  - **Techniques d'estimation dans les petites régions**

Estimation d'une statistique

- Le processus d'estimation établit un rapport entre l'échantillon et la population (inférence)
  - Fait via l'utilisation des poids d'échantillonage
- Que sont les poids d'échantillonnage?
  - Nombre de personnes que la personne interviewée représente dans la population
    - Exemple: poids = 500

Estimation d'une statistique

- **Que sont les poids d'échantillonnage? (suite)**
  - **Basés sur la probabilité de sélection**
    - **Une personne sélectionnée selon une fraction de sondage de 1% aura un poids (initial) de 100**
    - **Les fractions de sondage diffèrent d'une région à une autre, conséquemment, les poids diffèrent d'une personne à l'autre**
  - **Corrigés pour la non-réponse totale, et pour être en accord avec les estimations de population**

Estimation d'une statistique

- **Nom des variables de poids:**

Enquête	Cycle	Fichier	Détails	Nom (Maître/Partagé)
ENSP	1	Général		WT54 / SHRWT5
		Santé		WT64 / SHRWT6
ENSP	2	Général		WT56 / WT56_S
		Santé	Sauf HPS et Services santé enfants	WT66 / WT66_S
ENSP	3	Général		WT58 / WT58_S
		Santé		WT68 / WT68_S
ESCC	1.1	HSI	Échantillon total	WTSA_M / WTSA_S
			Trimestre 4 seulement	WTSA_Q4M / WTSA_Q4S
			Achat échantillon Î.-P.-É.	WTSA_PEM / WTSA_PES
			C.-B. - 16 régions	WTSAM
ESCC	1.2	HSI	Échantillon total	WTSB_M / WTSB_S

Estimation d'une statistique

- **Comment incorporer les poids dans les calculs:**
  - **En SPSS:**
    - **Utiliser "Weight Cases" sous le menu "Data"**
  - **En SAS:**
    - **Utiliser l'énoncé WEIGHT à l'intérieur de la procédure utilisée pour calculer la statistique**
- **Exemple:**
  - **Estimation du nombre de diabétiques au Québec**

Estimation d'une statistique

- **Exercice:**
  - **Estimer le nombre de diabétiques dans la région choisie (nombre total et proportion)**
  - **Comparer les proportions pondérées et non-pondérées**

Estimation de la précision

- Précision = Erreur d'échantillonnage
- Du fait que les résultats sont obtenus d'un échantillon et non d'un recensement
- Précision est fonction de:
  - Taille de l'échantillon et la population
  - Plan d'échantillonnage utilisé (effet du plan)
  - Magnitude de la proportion estimée

Estimation de la précision

- **Mesures de précision:**
  - **Variance, Écart-type, Intervalle de confiance**
  - **Coefficient de variation (CV)**
    - **CV = Écart-type de l'estimation x 100%**
      Estimation
    - **E.g.: 24% de la pop. sont des fumeurs réguliers,**
      Écart-type = 0.003
      CV = 0.003 / 0.24 x 100% = 1.25%
    - **CV permet la comparaison de la précision d'estimations d'échelles différentes**

Estimation de la précision

- **Calcul d'une estimation est simple (utiliser le poids)**
- **Lorsqu'on utilise les données de l'ENSP/ESCC, la précision d'une estimation est plus complexe à calculer**
  - **Pourquoi?**
    - **Données sont recueillies à l'aide d'un plan de sondage complexe**
    - **Pour les plans de sondage complexes, il n'existe pas de formule directe pour la calcul de la précision**

Plan de sondage complexe

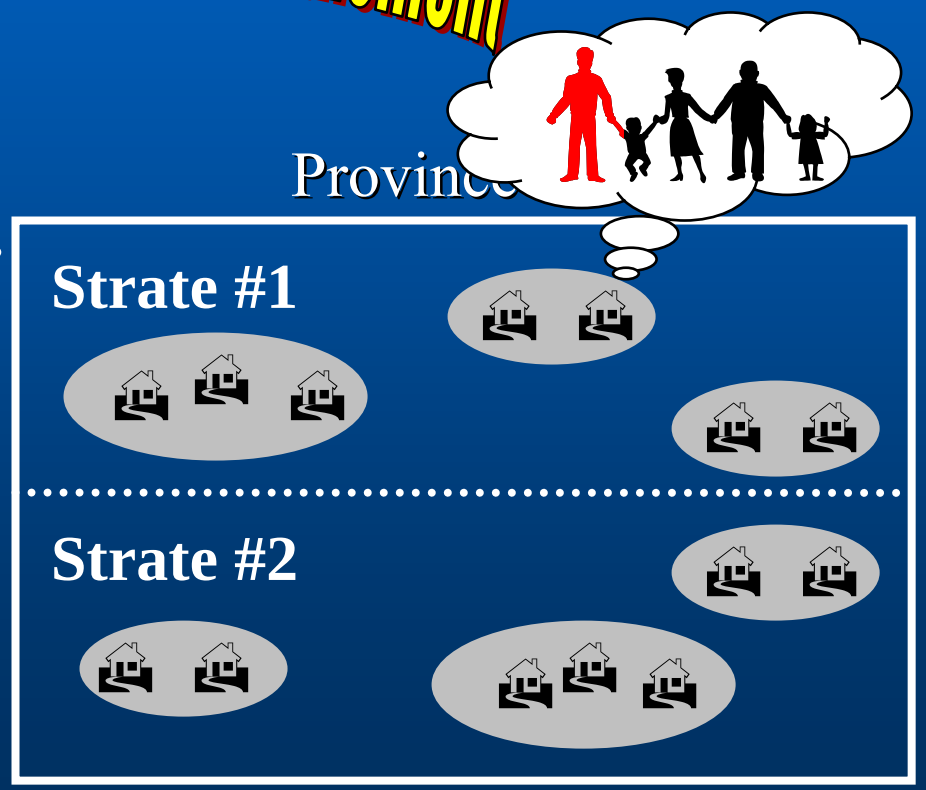
Illustration seulement

#1: La province est divisée en <u>strates</u>

#2: Des grappes sont sélectionnées à l'intérieur des strates (échant. PPT) (<u>1er degré</u>)

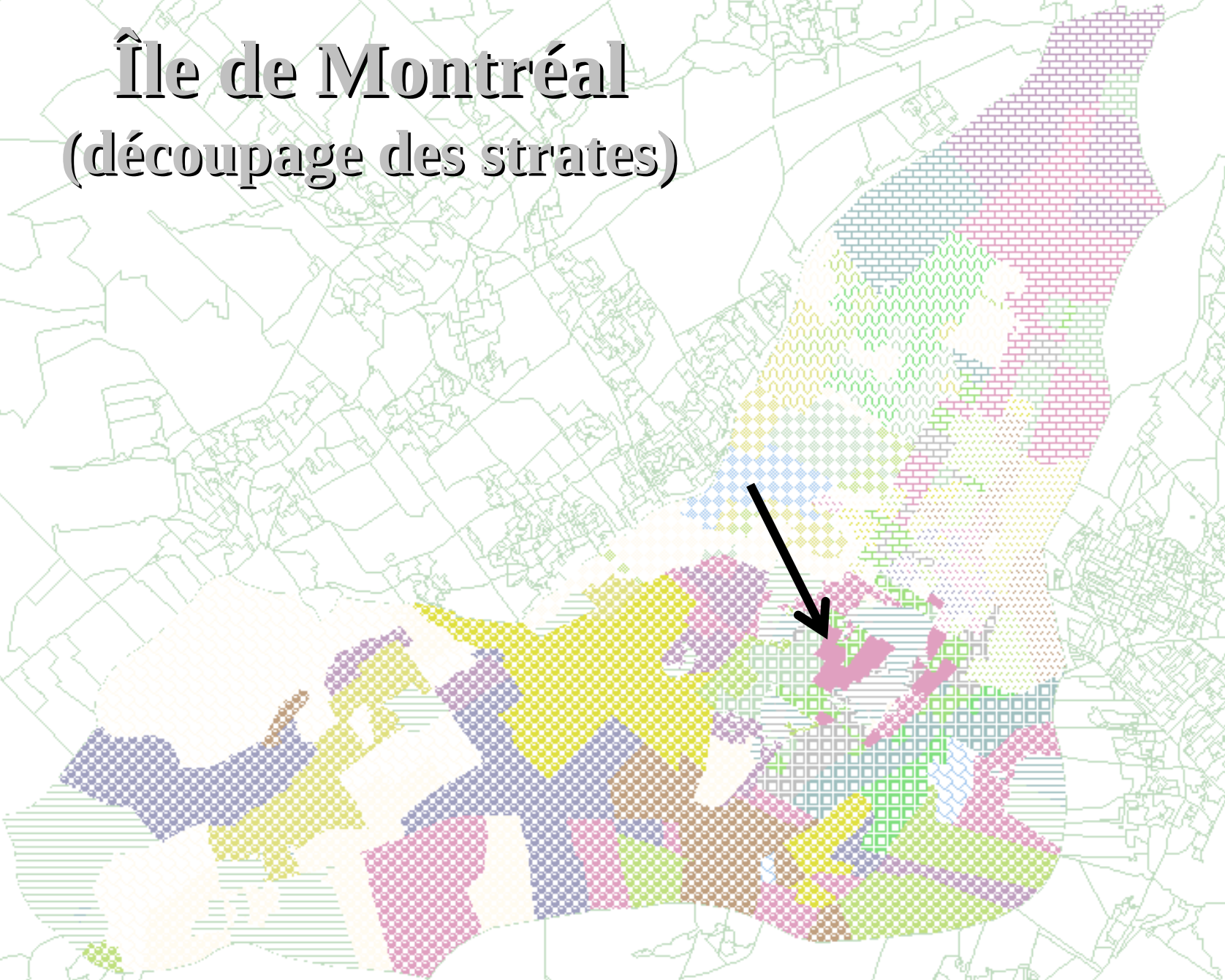
#3: Logements sélectionnés à l'intérieur des grappes (<u>2ème degré</u>)

#4: Personnes sélectionnées à l'intérieur des logements répondants (<u>3ème degré</u>)

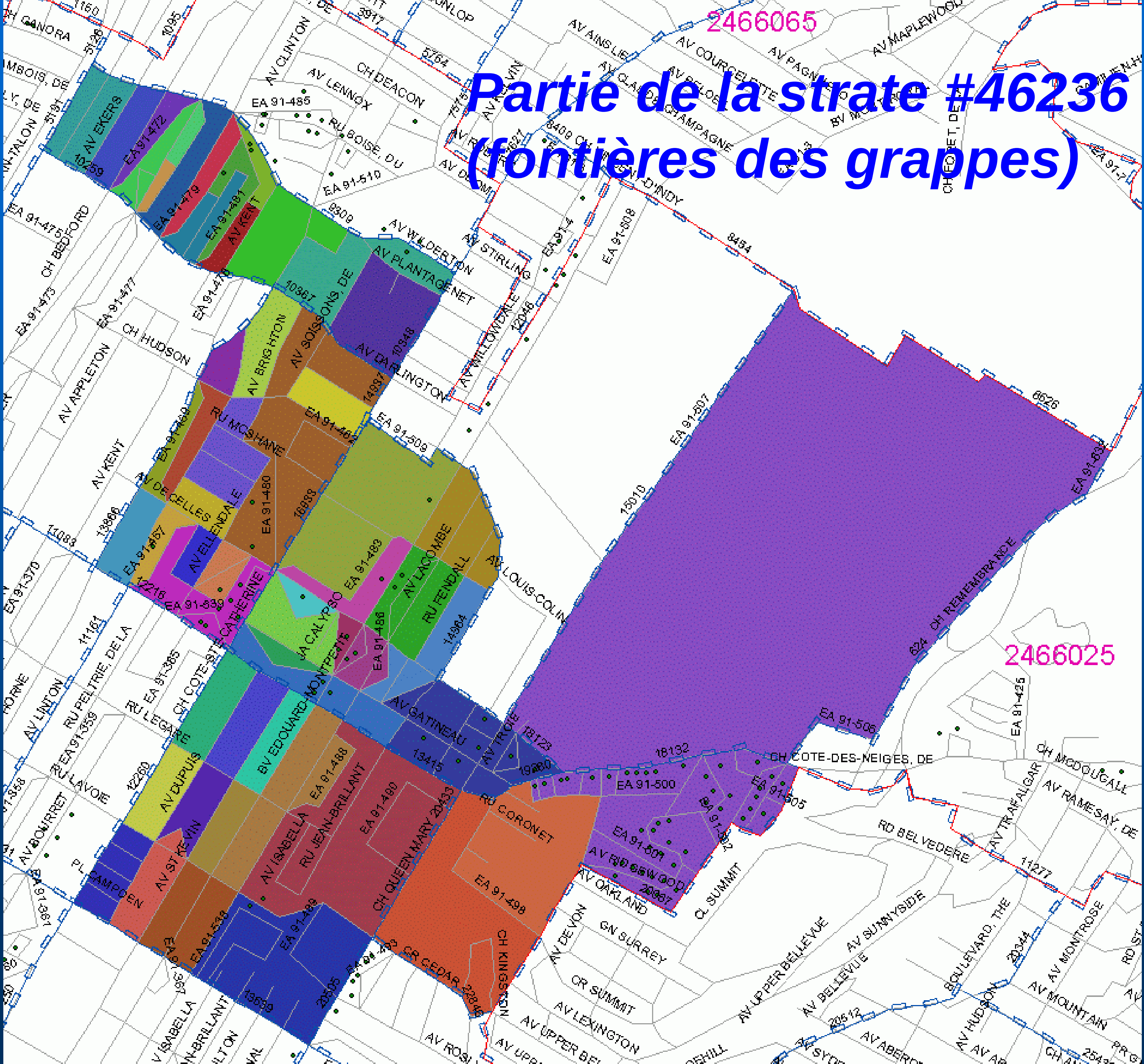


Île de Montréal

(découpage des strates)

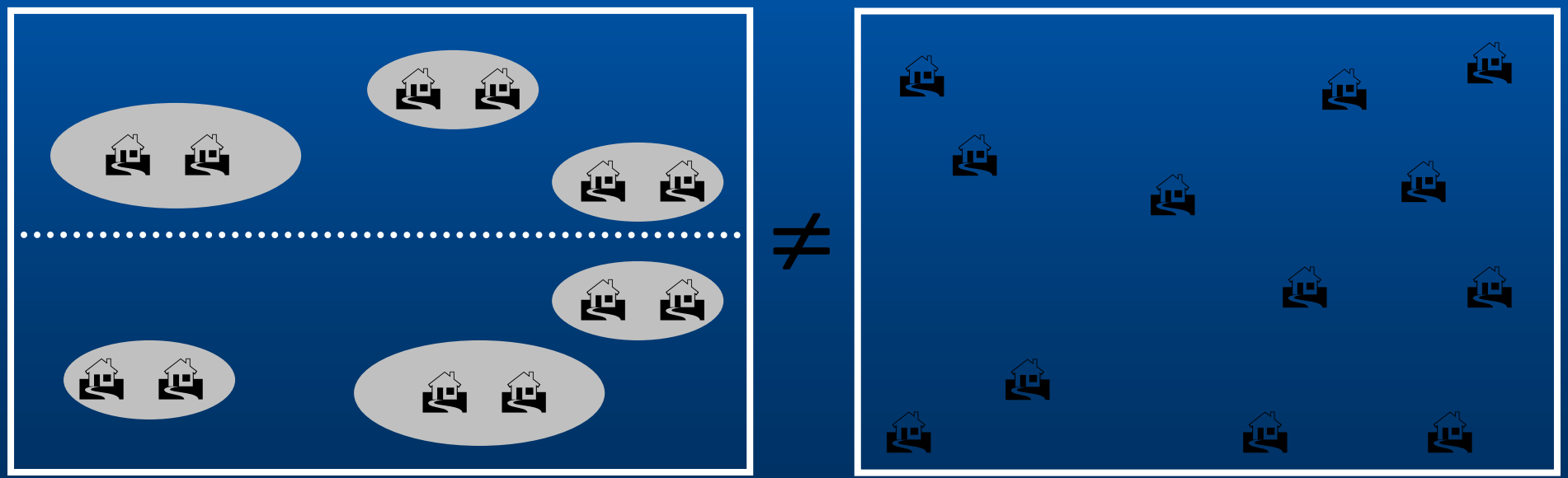


Partie de la strate #46236 (fontières des grappes)



Plan de sondage complexe

plan complexe ≠ plan aléatoire simple



Plan de sondage complexe

- **Pourquoi utiliser un tel plan?**
  - **Meilleur couverture de la région d'intérêt (stratification)**
  - **Efficace pour la collecte; moins de déplacement, moins coûteux (mise en grappes)**

  Problème: Estimation de la précision est plus complexe

Estimation de la précision

- **Comment le plan de sondage complexe affecte-il la précision des estimations?**
  - **Stratification réduit la variabilité (plus de précision)**
  - **Mise en grappes augmente la variabilité (moins de précision)**
  - **Globalement, le plan complexe à plusieurs degrés a l'effet d'augmenter la variabilité (moins précis qu'un échantillonnage aléatoire simple)**

Estimation de la précision

- **Pour l'estimation de la précision avec un plan de sondage complexe à plusieurs degrés:**
  - **Il n'existe pas de formule d'estimation de la variance; on a recours à des méthodes approximatives**
  - **NOTE: tenir compte du plan lors du calcul de la précision est aussi crucial que d'utiliser les poids d'échantillonnage lors de l'estimation de la statistique**

Estimation de la précision

- **Méthodes approximatives pour le calcul de la précision:**
  - **Linéarisation de Taylor**
  - **Méthodes par replication:**
    - **Répliques équilibrées répétées (BRR)**
    - **Jackknife**
    - **Bootstrap**

Estimation de la précision

- **Méthodes par répliques (Motivation):**
  - On peut estimer la variance de l'estimation d'un paramètre en utilisant un grand nombre de sous-échantillons différents tirés de l'échantillon original
    - **Chaque sous-échantillon, appelé réplique, est utilisé pour estimer le paramètre**
    - **La variabilité entre les estimations résultantes est utilisée pour estimer la variance de l'estimation de l'échantillon complet**
  - **Les méthodes par répliques diffèrent dans leurs façons de construire les répliques**

Estimation de la précision

- **Principe des méthodes par répliques:**
  - **Vous voulez estimer la précision de votre estimation du nombre de fumeurs au Canada**
  - **Prenez un grand nombre de sous-échantillons de votre échantillon**
  - **Pour chaque sous-échantillon, calculer l'estimation du nombre de fumeurs**
  - **Pour estimer la variance, calculer la variance entre les estimations provenant des sous-échantillons**

Estimation de la précision

- **Méthode du bootstrap:**
  - **Présenté dans le cadre d'un atelier pratique**
- **Le résultat est un fichier contenant 500 poids bootstrap (représentant les 500 répliques bootstrap)**
  - **Utilisé pour calculer la précision (variance, CV)**
  - **Programme Bootvar**
  - **Information confidentielle**
    - **Avec les fichiers maître & partagé seulement**
  - **Poids bootstrap coordonnés pour l'ENSP**

▶ Exemple avec le Bootvar pour le diabète

Obs	GEOA_HR4	type	var1	var2	yhat	bs_sd	bs_cv	cil95	ciu95
1	2401	Total	diab	Aucune	11197.41	1602.82	14.31	8055.88	14338.94
2	2401	Rapport	diab	total	6.42	0.92	14.31	4.62	8.22
3	2402	Total	diab	Aucune	10456.54	3609.92	34.52	3381.10	17531.98
4	2402	Rapport	diab	total	4.33	1.49	34.52	1.40	7.26
5	2403	Total	diab	Aucune	26320.72	4847.27	18.42	16820.07	35821.37
6	2403	Rapport	diab	total	4.73	0.87	18.42	3.02	6.44
7	2404	Total	diab	Aucune	16485.61	3504.46	21.26	9616.86	23354.36
8	2404	Rapport	diab	total	4.04	0.86	21.26	2.35	5.72
9	2405	Total	diab	Aucune	10249.12	3116.63	30.41	4140.53	16357.71
10	2405	Rapport	diab	total	4.19	1.27	30.41	1.69	6.69
11	2406	Total	diab	Aucune	78298.47	11005.69	14.06	56727.31	99869.63
12	2406	Rapport	diab	total	4.99	0.70	14.06	3.61	6.36
13	2407	Total	diab	Aucune	11739.11	2477.00	21.10	6884.20	16594.02
14	2407	Rapport	diab	total	4.38	0.92	21.10	2.57	6.20
15	2408	Total	diab	Aucune	5205.24	926.24	17.79	3389.81	7020.67
16	2408	Rapport	diab	total	4.20	0.75	17.79	2.74	5.67
17	2409	Total	diab	Aucune	3756.61	916.56	24.40	1960.16	5553.06
18	2409	Rapport	diab	total	4.83	1.18	24.40	2.52	7.14
19	2410	Total	diab	Aucune	290.46	111.48	38.38	71.95	508.97
20	2410	Rapport	diab	total	2.02	0.78	38.38	0.50	3.54
21	2411	Total	diab	Aucune	2357.23	703.32	29.84	978.73	3735.73
22	2411	Rapport	diab	total	2.76	0.82	29.84	1.15	4.38
23	2412	Total	diab	Aucune	15965.09	2549.22	15.97	10968.62	20961.56
24	2412	Rapport	diab	total	4.82	0.77	15.97	3.31	6.33
25	2413	Total	diab	Aucune	9386.64	2315.15	24.66	4848.95	13924.33
26	2413	Rapport	diab	total	3.16	0.78	24.66	1.63	4.69
27	2414	Total	diab	Aucune	17490.30	2817.56	16.11	11967.89	23012.71
28	2414	Rapport	diab	total	5.29	0.85	16.11	3.62	6.96
29	2415	Total	diab	Aucune	20752.31	4780.71	23.04	11382.11	30122.51
30	2415	Rapport	diab	total	5.26	1.21	23.04	2.88	7.63
31	2416	Total	diab	Aucune	55656.67	9335.06	16.77	37359.96	73953.38
32	2416	Rapport	diab	total	5.00	0.84	16.77	3.35	6.64

Estimation de la précision

- **Utilisation des poids d'échantillonnage et poids bootstrap**

Estimation pour le diabète - Canada ÉCARTS-TYPES	% diabétique	
	Estimation	É.-Type
Non-pondérée	4.1	0.162
Pondérée	3.5	0.151
Bootstrap	3.5	0.177
Source: ENSP - Fichier santé maître 1998		

Estimation de la précision

- Alternative au bootstrap: tableaux de CV
- Que sont-ils?
  - Tableaux de variabilité échantillonnale approximative
  - Produit à l'échelle du Canada (total & groupes d'âge), par province, et par région sous-provinciale (lorsqu'applicable)
  - Utile pour les estimations d'un total (d'une variable catégorique) ou d'une proportion

Enquête sur la santé dans les collectivités canadiennes - 2000/2001

Tableaux de la variabilité d'échantillonnage approximative : Région de Montréal-Centre (24906)

NUMÉRATEUR DU POURCENTAGE ('000)	POURCENTAGE ESTIMÉ												
	0.1%	1.0%	2.0%	5.0%	10.0%	15.0%	20.0%	25.0%	30.0%	35.0%	40.0%	50.0%	70.0%
1	91.1	90.7	90.2	88.8	86.5	84.0	81.5	78.9	76.2	73.5	70.6	64.4	49.9
2	********	64.1	63.8	62.8	61.1	59.4	57.6	55.8	53.9	52.0	49.9	45.6	35.3
3	********	52.4	52.1	51.3	49.9	48.5	47.1	45.6	44.0	42.4	40.8	37.2	28.8
4	********	45.3	45.1	44.4	43.2	42.0	40.8	39.5	38.1	36.7	35.3	32.2	25.0
5	********	40.6	40.3	39.7	38.7	37.6	36.5	35.3	34.1	32.9	31.6	28.8	22.3
. . . . .													
10	********	28.7	28.5	28.1	27.3	26.6	25.8	25.0	24.1	23.2	22.3	20.4	15.8
. . . . .													
15	********	23.4	23.3	22.9	22.3	21.7	21.0	20.4	19.7	19.0	18.2	16.6	12.9
. . . . .													
20	**********	20.2	19.9	19.3	18.8	18.2	17.6	17.0	16.4	15.8	14.4	11.2	
21	**********	19.7	19.4	18.9	18.3	17.8	17.2	16.6	16.0	15.4	14.1	10.9	
22	**********	19.2	18.9	18.4	17.9	17.4	16.8	16.3	15.7	15.1	13.7	10.6	
23	**********	18.8	18.5	18.0	17.5	17.0	16.5	15.9	15.3	14.7	13.4	10.4	
24	**********	18.4	18.1	17.6	17.2	16.6	16.1	15.6	15.0	14.4	13.2	10.2	
25	**********	18.0	17.8	17.3	16.8	16.3	15.8	15.2	14.7	14.1	12.9	10.0	
30	**********	16.5	16.2	15.8	15.3	14.9	14.4	13.9	13.4	12.9	11.8	9.1	
35	**********	15.0	14.6	14.2	13.8	13.3	12.9	12.4	11.9	10.9	8.4		
40	**********	14.0	13.7	13.3	12.9	12.5	12.1	11.6	11.2	10.2	7.9		
45	**********	13.2	12.9	12.5	12.2	11.8	11.4	11.0	10.5	9.6	7.4		
50	**********	12.6	12.2	11.9	11.5	11.2	10.8	10.4	10.0	9.1	7.1		
55	**********	12.0	11.7	11.3	11.0	10.6	10.3	9.9	9.5	8.7	6.7		
60	**********	11.5	11.2	10.8	10.5	10.2	9.8	9.5	9.1	8.3	6.4		
65	**********	11.0	10.7	10.4	10.1	9.8	9.5	9.1	8.8	8.0	6.2		
70	**********	10.6	10.3	10.0	9.7	9.4	9.1	8.8	8.4	7.7	6.0		
75	**********	10.3	10.0	9.7	9.4	9.1	8.8	8.5	8.2	7.4	5.8		
80	**********	9.7	9.4	9.1	8.8	8.5	8.2	7.9	7.2	5.6			
85	**********	9.4	9.1	8.8	8.6	8.3	8.0	7.7	7.0	5.4			
90	**********	9.1	8.9	8.6	8.3	8.0	7.7	7.4	6.8	5.3			
95	**********	8.9	8.6	8.4	8.1	7.8	7.5	7.2	6.6	5.1			
100	**********	8.6	8.4	8.2	7.9	7.6	7.3	7.1	6.4	5.0			
105	**********	8.4	8.2	8.0	7.7	7.4	7.1	6.9	6.2	4.9			

Estimation de la précision

- **Lignes directrices concernant la variabilité échantillonnale**

Type d'estimation	CV	Lignes directrices
Acceptable	0.0 - 16.5	Diffusion sans contrainte
Marginale	16.6 - 33.3	Diffusion sans contrainte mais avec mise en garde aux utilisateurs de la haute variabilité de l'échantillonnage (identifié par E)
Inacceptable	> 33.3	Pas de publication (identifiée par F)

Estimation de la précision

- **Exercice: Estimation du CV pour la proportion de diabétiques**
  - **Le tableau de CV peut être utilisé pour vérifier si on a assez d'échantillon avant de se lancer dans l'analyse**

Module 3

Comment traiter les données manquantes

Aperçu

- Qu’est-ce qu’une donnée manquante
- Types de données manquantes dans les enquêtes de Statistique Canada
- Comment traiter les données manquantes
- Données manquantes avec logiciels statistiques

Données manquantes

- **Les données manquantes sont dûes à la non-réponse à certaines ou à toutes les questions**
  - **Exemples de non-réponse:**
    - **Refus**
    - **Ne sait pas**
    - **Sauts dans le questionnaire**
    - **Mort d'un répondant longitudinal**

Non-réponse

- 2 types de non-réponse:
  - Non-réponse totale (par unité)
    - Aucune information recueillie
  - Non-réponse partielle (par item)
    - Certaines variables recueillies

Non-réponse

- **Non-réponse totale:**
  - **Toutes les variables sont manquantes pour le répondant**
    - **Exemples:**
      - **Refus complet**
      - **Incapable de contacter le répondant**
      - **Respondant absent pour la durée de l'enquête (si aucun répondant par procuration permis ou disponible)**
      - **Problème de langue**

Non-réponse

- **Non-réponse totale:**
  - **Prise en considération en ajustant les poids d'échantillonnage**
    - **Fait par les méthodologistes de Statistique Canada**
      - **En utilisant des classes d'ajustement pour éliminer un potentiel biais causé par la non-réponse**
    - **Les utilisateurs n'ont pas à modifier les données**

Non-réponse

- **Non-réponse partielle:**
  - **Certaines variables sont manquante pour le répondant**
    - **Exemples:**
      - **Répondant refuse de répondre à certaines questions spécifiques**
      - **Répondant ne connait pas la réponse**
      - **Données non-disponibles**
  - **Il y a différentes approches pour traiter la non-réponse partielle**

Types de données manquantes dans les enquêtes de Statistique Canada

- Il n'y a pas de « trous » dans les fichiers de données de Statistique Canada. Une valeur spécifique identifie les données manquantes

Type de données manquantes	Valeur
Sans objet	6, 96, 996
Ne sait pas	7, 97, 997
Refus	8, 98, 998
Non déclaré	9, 99, 999

- Note: pour l'ENSP (longitudinal) les morts sont codés à "Non déclaré"

Comment traiter les données manquantes

- **Dépend du type de donnée manquante, du type d'analyse et du nombre de données manquantes**
  - **1: Conserver les données manquantes et les rapporter séparément**
  - **2: si données manquantes aléatoires: enlever les enregistrements avec données manquantes (repondérer pour les estimations de totaux)**
  - **3: si données manquantes non-aléatoires: enlever les enregistrements avec données manquantes et repondérer**
  - **4: Imputation: remplacer chaque valeur manquante par une valeur de remplacement)**

Comment traiter les données manquantes

- **Exemples d'imputation:**
  - **déterministe**
  - **par donneur**
  - **Par la moyenne**
  - **historique (si longitudinal)**

- **Une analyse devrait être faite pour vérifier la validité de l'imputation (ex: taux d'imputation, etc.)**
- **L'imputation doit être mentionnée dans le rapport final**

Comment traiter les données manquantes

- **Exemples: Nombre de nuits passées comme patient**

Répondant	A passé une nuit	# de nuits	# de nuits
1	Oui	2	2
2	Non	996	0
3	Oui	15	15
4	Non	996	0
5	Non	996	0
6	Oui	10	10

Comment traiter les données manquantes

- **Exemples:**

Répondant	Fumeur	Sexe	Poids
1	Oui	F	12,5
2	Non	F	12,5
3	8	F	
4	Non	F	12,5
5	Non	F	12,5
6	8	H	
7	Oui	H	25
8	8	H	
9	8	H	
10	Non	H	25

Problème potentiel:

Conclusions:

Non-réponse non-aléatoire (les
mmes ont tendance à refuser
vantage que les femmes)

Solution possible:

Ajuster les poids

Dans la population:

37.5% fumeurs

62.5% non-fumeurs

Dans l'échantillon:

25% fumeurs

75% non-fumeurs

(non-réponse partielle ignorée)

66.6% non-fumeurs

(non-réponse partielle ignorée)

Données manquantes avec logiciels statistiques

- **Si rien n'est fait, les logiciels considèrent les valeurs manquantes (6, 7, 8, 9, etc…) comme des valeurs possibles**
  - **SAS: Les utilisateurs doivent recoder la variable (avec ' ' ou .)**
  - **SPSS: Les utilisateurs peuvent recoder la variable (avec ' ' ou .) ou utiliser l'énoncé "MISSING VALUES"**
    - **Note: Pour certaines procédures (ex: regression), l'enregistrement complet est rejeté lorsqu'au moins une variable d'intérêt est manquante**

Exemple

- **Objectif: Créer une variable catégorique pour l'indice de masse corporelle (HWTADBMI) et recoder les valeurs manquantes:**

  - **Les catégories sont:**

    - **0-20 = 1 - bas**
    - **20-30 = 2 - moyen**
    - **30 et + = 3 - élevé**

  - **Les valeurs manquantes sont:**

    - **999.6 = sans objet** (Univers: Répondants âgés de 20 à 64 ans qui ont répondu MAMA_037 <> 1)
    - **999.9 = non déclaré**

Exercice

- **Objectif: Seulement pour votre région de santé, estimer le ratio et le nombre des personnes (dans la population) avec un revenu total moyen de moins de 20 000$...**
  - **1: … en conservant les données originales (en rapportant les valeurs manquantes)**
  - **2: … en recodant les valeurs 7, 8 et 9 à “valeur manquante” (ignorant ainsi les valeurs manquantes)**
    - **Variable: INCA_3A:**

      Revenu total du ménage - < 20 000$ ou >= 20 000$

      1- Moins de 20 000$
      2- 20 000$ ou plus
      3- Aucun revenu
      7- Ne sait pas
      8- Refus
      9- Non déclaré

Exercice

- **1: Pour le Québec, en rapportant les valeurs manquantes:**

INCA_3A	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	1054358	16.92	1054358	16.92
2	4838040	77.64	5892398	94.56
3	21364.97	0.34	5913763	94.91
7	102719.1	1.65	6016482	96.56
8	116589.1	1.87	6133071	98.43
9	98034.67	1.57	6231106	100.00

Exercice

- **2: Pour le Québec, en ignorant les valeurs manquantes:**

INCA_3A	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	1075723	18.19	1075723	18.19
2	4838040	81.81	5913763	100.00

Exercice

- **Ratio:**
  - **En rapportant les valeurs manquantes:**

    17,3% de la population a un revenu du ménage < $20,000

    77,6% de la population a un revenu du ménage >= $20,000

    5,1% de la population ne sait pas ou refuse de répondre

  - **En ignorant les valeurs manquantes:**

    18,2% de la population a un revenu du ménage < $20,000

    81,8% de la population a un revenu du ménage >= $20,000

    La non-réponse partielle a été ignorée (assumant une non-réponse aléatoire)

Exercice

- **Totals:**
  - **En rapportant les valeurs manquantes:**

    Dans la population:

    1 075 723 personnes ont un revenu du ménage < $20,000

    4 838 040 personnes ont un revenu du ménage >= $20,000

    317 342 personne ne savent pas ou refusent de répondre

  - **En ignorant les valeurs manquantes:**

    6 231 106 X 18,2% = 1 134 061

    6 231 106 X 81.8% = 5 097 044

    Dans la population:

    1 134 061 personnes ont un revenu du ménage < $20,000

    5 097 044 personnes ont un revenu du ménage >= $20,000

    La non-réponse partielle a été ignorée (assumant une non-réponse aléatoire)

Module 4

Survol de quelques méthodes d’analyse transversale

Aperçu

- Prévalence (estimation & tests de différence)
- Régression linéaire et logistique
- Standardisation âge-sexe (pour des comparaisons de populations)
- Avertissements concernant la comparaison d’estimations transversales dans le temps

Analyse transversale

- **Analyses possibles:**
  - **Statistiques descriptives (totaux, rapports, moyennes):**
    - **Prévalence: Le nombre de cas d'une certaine maladie ou autre condition dans une population donnée à un temps donné.**
  - **Étudier la relation entre deux ou plusieurs variables**
    - **Modèles de régression linéaire**
    - **Modèles de régression logistique**
    - **etc...**
    - **Le type d'analyse dépend du type de variables d'intérêt (qualitative vs quantitative)**

Analyse simple

- **Statistiques descriptives: prévalence (totaux, moyennes, rapports)**

  Étape 1: Identifier la (les) variable(s)

  Étape 2: Recoder les valeurs manquantes (si désiré)

  Étape 3: Utiliser les poids

  Étape 4: Calculer l'(les) estimation(s)

  Étape 5: Calculer la (les) variance(s)

Analyse simple: exemple

- **Étape 1: Identifier la variable**
  - **CCCA_91B: Souffre d'emphysème / broncho-pneumopathie chronique obstructive (BPCO)**
    - **(question posée aux 30 ans et plus seulement)**

Source: ESCC cycle 1.1 – fichier bidon

Contenu	Code	Échantillon	Population
OUI	1	118	57 603
NON	2	8 298	4 464 320
SANS OBJET	6	2 832	1 703 636
NE SAIT PAS	7	2	2 285
NON DÉCLARÉ	9	5	3 261
	TOTAL	11 255	6 231 106

Analyse simple: exemple

- **Étape 2: Recoder les valeurs manquantes**

	Échantillon		Échantillon
OUI (1)	118	→	118
NON (2)	8 298	→	8 298
SANS OBJET (6)	2 832		.
NE SAIT PAS (7)	2		.
NON DÉCLARÉ (9)	5		.
TOTAL	11 255		8 416

Analyse simple: exemple

- **Étape 3: Utiliser les poids**

	Échantillon
OUI (1)	118
NON (2)	8 298
SANS OBJET (6)	2 832
NE SAIT PAS (7)	2
NON DÉCLARÉ (9)	5
TOTAL	11 255

Échantillon	Population
118	57 603
8 298	4 464 320
.	
.	
.	
8 416	4 521 923

Analyse simple: exemple

- **Step 4: Calculer les estimations**
  - **exemples:**

	Population
OUI (1)	57 603
NON (2)	4 464 320
SANS OBJET (6)	
NE SAIT PAS (7)	2285
NON DÉCLARÉ (9)	3261
TOTAL	4 521 923

- 1,3% des personnes de 30 ans ou plus ont l'emphysème ou une BPCO (ignorant la non-réponse partielle) **(57 603 ÷ 4 521 923) X 100**

- 57 675 personnes de 30 ans ou plus ont l'emphysème ou une BPCO (ignorant la non-réponse partielle) (hypothèse: 1,3% des non-répondants ont l'emphysème ou une BPCO)

Analyse simple: exemple

- **Step 5: Calculer les variances**
  - **Le programme Bootvar (fourni avec les données) permet d'obtenir l'écrat-type, le CV et un intervalle de confiance**
  - **Exercice: Estimer la variance à l'aide du programme:**

\\ciqss-s2\utilisateurs$\formation\ordinateurXX\Bootvarf_BPCO.sas

où XX est le numéro de chaque ordinateur

Difference de rapports

- **Comparaison de populations:**
  - **On désire tester si le taux des hommes souffrant d'emphysème ou de BPCO est significativement différent du taux des femmes**

	Hommes	Femmes
Oui	36 152	21 451
Non	2 152 818	2 311 502

- **Hommes: 1,65%** des hommes de 30 ans ou plus souffrent d'emphysème ou de BPCO
- **Femmes: 0,92 %** des femmes de 30 ans ou plus souffrent d'emphysème ou de BPCO

Difference de rapports

- **Test statistique: Test de Student (test T)**

taux_h = proportion des hommes = 1,65%
taux_f = proportion des femmes = 0,92%

Test d'hypothèse: **H_0: taux_h = taux_f**
H_1: taux_f ≠ taux_f

Statistique:

$$Z = \frac{(taux\_h - taux\_f)}{sd\ (taux\_h - taux\_f)}$$

Test: **Au niveau ✂ = 0,05, on ne rejette pas H_0 si** | z | <= 1,96
Sinon, on rejette H_0.

Resultats:

$$Z = \frac{(1,65\% - 0,92\%)}{sd\ (taux\_h - taux\_f)} = \frac{(1,65\% - 0,92\%)}{0,40} = 1,83$$

(peut-être obtenu avec Bootvar)

Analyse plus complexe

- **Modèles de régression**
  - **La régression est une méthode statistique qui utilise la relation entre deux ou plusieurs variables de telle sorte qu'une variable peut être prédite à partir des autres.**

    (Applied linear statistical models, Neter & al.)

  - **La régression linéaire consiste en un ensemble de techniques utilisées pour explorer la relation entre des variables.**

    (Applied Linear Regression, Weisberg)

Analyse plus complexe

- **Modèle de régression linéaire:**

$$Y = \alpha + \beta_1 * X_1 + \beta_2 * X_2 + \ldots + \beta_j * X_j + \varepsilon$$

  - Y est quantitative
  - X_i peut être quantitative ou qualitative (catégorique) (les variables catégoriques doivent être «dichotomisées»)

Analyse plus complexe

- **Exemple d'un modèle de régression linéaire:**
  - **IMC (indice de masse corporelle) VS sexe et âge**

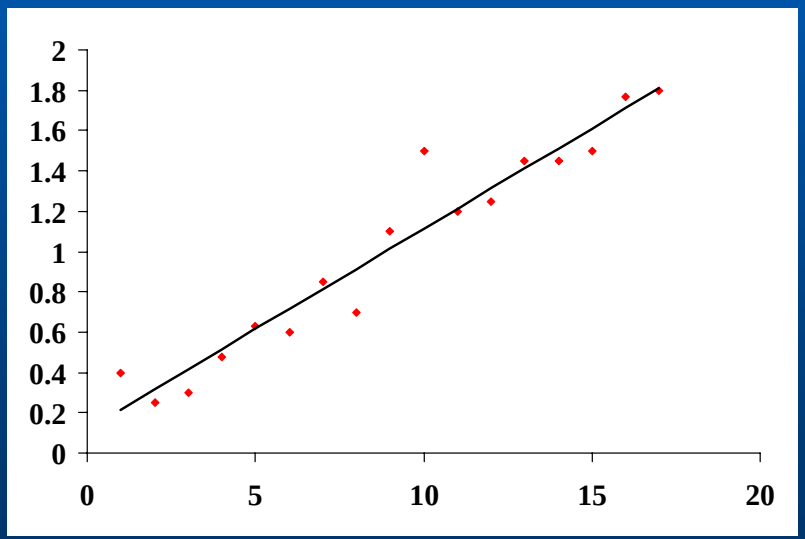
$$IMC = \alpha + \beta_1 * FEMME + \beta_2 * AGE + \varepsilon$$

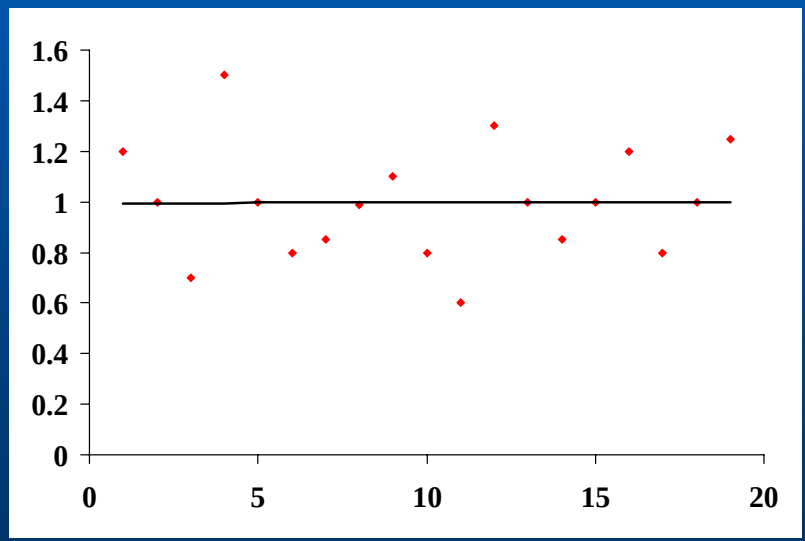
    - **Les variables catégoriques doivent être "dichotomisées»**
      (If SEX=FEMME then FEMME=1; else FEMME=0)
    - **AGE est une variable quantitative**

Analyse plus complexe

- **Illustration d'un modèle de régression linéaire univariée:**

Modèle: $Y = \alpha + \beta_1 * X_1 + \varepsilon$





L'objectif est d'estimer si la pente ($\beta 1$) et l'ordonné à l'origine (α) sont statistiquement différents de 0

Analyse plus complexe

- **Construire un modèle de régression:**
  - **1: Explorer les données**
  - **2: Choisir les variables indépendantes**
    - **utiliser ses connaissances et ses expertises**
    - **parsimonie**
    - **étudier la relation entre les variables**
  - **3: Modèles préliminaires**
    - **Vérifier que toutes hypothèses sous-jacentes sont respectées**
    - **Diagnostiques (transformation, valeurs extrêmes, etc.)**
  - **4: Rafinement du modèle et sélection du modèle final**

Analyse plus complexe: exemple

- **Exemple: Consommation de fruits et légumes**

  (Source: Rapports de la santé, *Vol 13 no.3*, Statistics Canada)

  - **Objectif:**

    Analyser les associations entre la fréquence de consommation de fruits et légumes et certains facteurs ou comportements reliés à la santé.

  - **Source des données:**

    Première moitié de l'ESCC, cycle 1.1 (collecté de sept. 2000 à fév. 2001)

  - **Technique d'analyse:**

    Régression linéaire multivariée

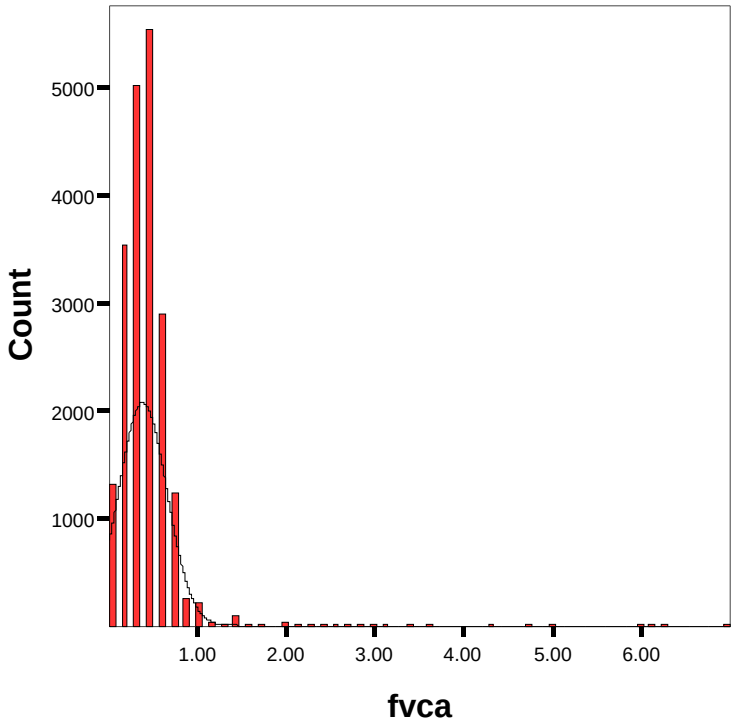
Analyse plus complexe: exemple

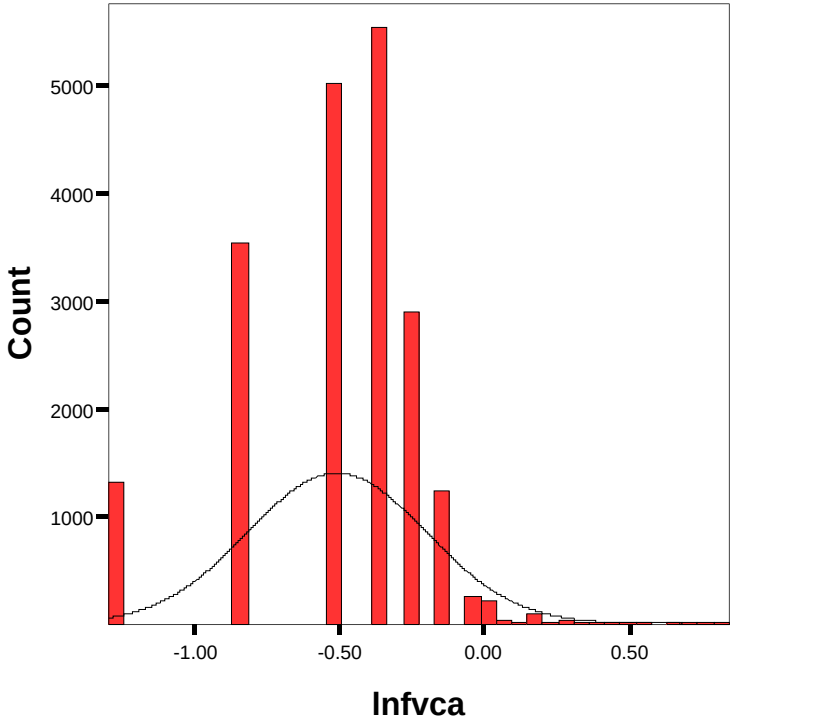
- **Préliminaire**
  - **Taux de réponse de la première moitié de l'ESCC: 80%, Non-réponse totale ajustée par Statcan (pondération)**
  - **Les valeurs manquantes n'ont pas été incluses dans l'analyse (ie: les enregistrements avec au moins une valeur manquante pour au moins une des variables utilisées dans le modèle ont été exclues de l'analyse)**
  - **Taux d'imputation: 7,6% (interviews par procuration)**
  - **Pour tenir compte du plan de sondage complexe de l'enquête, la méthode du bootstrap a été utilisée pour estimer la variance des coefficients de régression**

Analyse plus complexe: exemple

- **Variable dépendante: Fréquence quotidienne de consommation de fruits et légumes**
  - **Dérivée à partir des questions demandant combien de fois des fruits ou légumes spécifiques ont été consommés**
  - **Une transformation logarithmique de la variable dépendante a été faite en raison de l'asymétrie des données**

Analyse plus complexe: exemple





Analyse plus complexe: exemple

- **Varaibles prédicatrices (variables indépendantes):**
  - **Comportements et facteurs reliés à la santé**

- Activité physique
- Consommation de cigarette
- Indice de masse corporelle
- Dépendance à l'alcool
- Conditions chroniques
- Incapacité
- Évaluation personnelle de la santé
- Groupe d'âge
- Revenu du ménage
- Education
- État matrimonial

Analyse plus complexe: exemple

- **Résultats de la régression linéaire multivariée:**

Characteristiques	Beta
Acitvité physique	0.05 *
Cigarette	
▪ Non-fumeur	0.18 *
▪ Occasionnel	0.20 *
▪ quotidien *	--
Ind. de masse corp.	- 0.005 *
Dépendance à l'alcool	
▪ Non	0.08
▪ Oui *	--
Aucune cond. chron.	- 0.09 *

Characteristiques	Beta
Aucune incapacité	- 0.06
Perception de sa santé	0.04 *
Âge	0.01 *
Revenu du ménage	0.02 *
Education	0.02 *
État matrimonial	
▪ Marié/Con.de fait	0.13 *
▪ Célibataire	0.11 *
▪ Divorcé *	--
Ordonné	0.62

*** = valeur-p < 0.05**

Analyse plus complexe

- **Modèle de régression logistique:**
  - **Le modèle n'est pas linéaire, mais il peut être linéarisé à l'aide d'une transformation logarithmique. Le modèle devient:**

$$\log(f(Y)) = \text{intercept} + \beta_1 * X_1 + \beta_2 * X_2 + \ldots + \beta_j * X_j + \varepsilon$$

    - **Y doit être qualitative (catégorique) (souvent dichotomique)**
    - **X_i peut être quantitative ou qualitative**

Analyse plus complexe

- **Exemple d'un modèle de régression logistique:**
  - **Diabète vs sexe et âge**

DIAB = intercept + β_1*FEMME + β_2*AGE + ε

- **DIAB est une variable dichotomique (0 ou 1)**
- **Les variables catégoriques doivent être "dichotomisées»**
  (If SEX=FEMME then FEMME=1; else FEMME=0)

Analyse plus complexe

- **Exemple d'une sortie de Bootvar**

prov	beta	bhat	Odds	Wald	pvalue	bs_var	bs_sd	bs_cv	cil95	ciu95
10	Intercept	-4.00	0.01	255.38	0.0000	0.06	0.25	6.26	0.01	0.02
10	femme	0.85	2.35	11.81	0.0005	0.06	0.24	29.09	1.44	3.84
10	age	0.46	1.59	2.85	0.0908	0.07	0.27	59.14	0.92	2.73

Rapport de cotes (Odds): Rapport entre le risque d'une maladie ou d'un décès parmi les répondants exposés au facteur et le risque parmi les répondants qui ne sont pas exposés.

Analyse plus complexe: exemple

- **Exemple: Précarité alimentaire dans les ménages canadiens**

  (Source: Rapports de la santé, *Vol 12 no.4*, Statistics Canada)

  - **Objectif:**

    Déterminer la prévalence de la précarité alimentaire au Canada et les caractéristique des personnes les plus susceptibles de vivre dans ménage affecté par la précarité alimentaire

    Source des données:

    Composante transversale de l'ENSP, 1998-1999

    Technique d'analyse:

    Régression logistique multivariée

Analyse plus complexe: exemple

- **Préliminaire**
  - **Taux de réponse pour le cycle 3 de l'ENSP: 88,2%, Non-réponse totale ajustée par Statcan (pondération)**
  - **Les valeurs manquantes n'ont pas été incluses dans l'analyse (ie: les enregistrements avec au moins une valeur manquante pour au moins une des variables utilisées dans le modèle ont été excluses de l'analyse)**
  - **Pour tenir compte du plan de sondage complexe de l'enquête, la méthode du bootstrap a été utilisée pour estimer la variance des rapports de cotes**

Analyse plus complexe: exemple

- **Variable dépendante: Souffre de précarité alimentaire (oui/non)**
  - **Variables prédicatrices (variables indépendantes):**
    - Sexe
    - Groupe d'âge
    - Revenu du ménage
    - Source principale de revenu
    - Type de ménage
    - Propriétaire / locataire
    - État matrimonial
    - Statut d'immigrant
    - Autochtone (oui/non)

Analyse plus complexe: exemple

Characteristics	Odds Ratio
Sexe	
Hommes	1.06
Femmes *	1.00
Groupe d'âge	
0 - 17	4.82 *
18 - 44	4.22 *
45 - 64	2.71*
65 + *	1.00
Revenu du ménage	
Bas	7.96 *
Moyen	4.31 *
Élevé *	1.00
Source principale de revenu	
Salaires*	1.00
Compensations	1.71 *
Ass. sociale	3.06 *
Autres	1.02
Propriétaire / locataire	
Propriétaire*	1.00
Locataire	2.01 *

Characteristics	Odds Ratio
Type de ménage	
Couple avec enfants*	1.00
Couple sans enfant	0.98
Mère seule	1.41 *
Père seul	1.02
Personne seule	0.95
Autre	0.99
État matrimonial	
Marié *	1.00
Avec partenaire	1.06
Célibataire	0.79 *
Veuf	1.04
Divorcé	1.45 *
Statut d'immigrant	
Né au Canada *	1.00
Immigré depuis 0-9 ans	0.66 *
Immigré depuis +10 ans	1.05
Autochtone	
Oui	1.48 *
Non *	1.00

Standardisation

- **Quoi?**
  - **La standarisation est un technique utilisée pour comparer des sous-populations pour une variable donnée, alors qu'on contrôle pour la différence entre les profils de chaque sous-population par rapport à une autre variable**
- **Exemple:**
  - **La comparaison des taux d'asthme entre deux régions ayant des profils (distributions) âge-sexe différents**

Standardisation

- **Solution?**
  - **Appliquer le même profil (distribution) provenant d'une population standard, à toutes les sous-populations comparées**

Standardisation

- **Exemple:**

	RÉGION A		Pop std	RÉGION B	
Grp âge	**Pop%**	**Taux%**	**Pop%**	**Pop%**	**Taux%**
<60	~~80%~~	40%	50%	~~20%~~	40%
>=60	~~20%~~	60%	50%	~~80%~~	60%
Taux global:		~~44%~~			~~60%~~
Taux standardisé		50%			50%

Standardisation

- **Information requise:**
  - **La distribution par groupe âge-sexe pour la population standard (D1, D2, …, D10)**
  - **Taux pour la variable d'intérêt, calculé pour chaque groupe âge-sexe à l'intérieur de chaque région comparée (R1, R2, …, R10)**
- **L'équation!:**
  - **Taux std = (D1*R1)+(D2*R2)+….+(D10*R10)**

Standardisation

- Méthode directe (vs. indirecte)
- Les taux standardisés ne signifient rien en eux-mêmes, mais sont comparables
- L’estimation de la variance estimation pour ces taux standardisés requiert des modifications aux poids bootstrap (il faut les standardiser)

Avertissements concernant la comparaison d'estimations transversales dans le temps

- **Aspects qui peuvent expliquer les différences dans les estimations obtenues à partir de deux différentes enquêtes transversales**
  - **Aspects méthodologiques:**
    - **Population cible**
    - **Questionnaire**
    - **Collecte (période, taux de réponse, int. par procuration, etc…)**
    - **Pondération**
    - **Imputation**
    - **Méthode pour calculer la variance**
    - **Variabilité échantillonnale**
  - **Aspects contextuels :**
    - **Changements dans les normes relatives à la santé**
    - **Changement réel dans la population**

Module 6

Survol de logiciels pour l'analyse de données

Plan

- Points à vérifier lorsqu'on utilise un logiciel pour l'analyse de données
- Estimation de variance
- Télé-accès / support

Points à vérifier

- Traitement des valeurs manquantes (modèles multivariés)
- Utilisation adéquate des poids (pour l'estimation)
- Tenir compte du plan de sondage complexe dans l'estimation de la variance

Estimation de la variance

- Certains logiciels incluent des méthodes pour les plans de sondage complexes
  - Les méthodes supportées varient d'un logiciel à l'autre
  - Le bootstrap n'est pas disponible "directement"
    - Utiliser Bootvar

Complex Survey Design Variance Estimation for some Software Programs

	SAS	Stata	SUDAAN	WesVar	Bootvar
Approximative Methods Supported	Taylor	Taylor	Jackknife, BRR, Taylor, Bootstrap*	Jackkinfe, BRR, Bootstrap*	Bootstrap
Descriptive					
means	*surveymeans*	*svymean*	*descript*	yes	yes
totals	*surveymeans*	*svytotal*	*descript*	yes	yes
proportions	no	*svyprop*	*descript*	yes	yes
ratios	no	*svyratio*	*ratio*	yes	yes
crosstabulations	no	*svytab*	*crosstab*	yes	no
quantiles	no	no	*descript*	yes	no
Modelling					
linear regression	*surveyreg*	*svyreg*	*regress*	yes	yes
instrumental variable regression	no	*svyireg*	no	no	no
interval regression	no	*svyintrg*	no	no	no
logistic regression	no	*svylogit*	*logistic or rlogist*	yes	yes
probit regression	no	*svyprobt*	no	no	no
mutinomial logistic regression	no	*svymlog*	*multilog*	yes	no
ordered logistic regression	no	*svyolog*	*multilog*	no	no
ordered probit regression	no	*svyoprob*	no	no	no
poisson and log-linear regression	no	*svypois*	*loglink*	no	no
hierarchical linear models	no	no	no	no	no
proportional hazards models	no	no	*survival*	no	no

Télé-accès / support

- Le service de télé-accès pour l'ENSP & l'ESCC supporte principalement SAS & SPSS
- Dans les CDR, un grand éventail de logiciels sont disponibles

Module 7

Références

Références

- Rapport sur la santé
  - Disponible sur le site web de Statistique Canada (www.statcan.ca)
    - Nos produits et services
    - Payantes
    - Santé
    - Rapports sur la santé
- Litérature (livres, articles)
  - Voir document

Références

- **Cours - Statistique Canada**
  - **Méthodes statistiques pour l'analyse des données — Niveau intermédiaire (0428)**
    - **Lois de probabilité classiques et utilisation des tables statistiques**
    - **Estimation ponctuelle**
    - **Estimation par intervalle de confiance**
    - **Tests d'hypothèse et théorie de la décision statistique**
    - **Régression linéaire simple**
    - **Analyse des résidus**

Références

- **Cours – Statistique Canada**
  - **Analyse de données de survie (0409)**
    - **‘Accelerated failure time models’**
    - **Modèles de Weibull et Gompertz**
    - **Types de censure**
    - **Estimation du maximum de vraisemblance**
    - **Modèles à risques proportionnels**
    - **Risques concurrents**
    - **Covariables dépendant du temps**
    - **Analyse des données discrètes**
    - **Événements répétitifs**
    - **Censure à gauche, troncation à gauche**
    - **Etc.**

Références

- **Cours - Statistique Canada**
  - **Traitement de la non-réponse dans les enquêtes et recensements (0424)**
    - Concepts de base (définitions, niveaux de non-réponse, exemples)
    - Causes de la non-réponse
    - Méthodes permettant de réduire la non-réponse
    - Repondération
    - Imputation
    - Estimation de variance en présence d'imputation
    - Analyse sur données manquantes
    - Évaluation de la non-réponse

Références

- **Cours - Autres**
  - **Summer Programme in Data Analysis (SPIDA, Université York)**
    - **Série de cours intensifs (2 semaines) ayant pour but de former les chercheurs pour l'analyse de données porvenant de plans complexes, et spécifiquement les données longitudinales et multi-niveaux**
    - **www.math.yorku.ca/SCS/spida/home.html**
    - **Programme (2003):**
      - **Modèles linéaires; Régression logistique; Modèles linéaires généralisés; Modèles de régression non-paramétrique; Introduction aux modèles mixtes & hiérarchiques; etc.**

Références

- **Cours - Autres**
  - **ÉCOLE D'ÉTÉ DU CIED / CÉETUM / CIQSS**
    - **www.ciqss.umontreal.ca/activites_formation.htm**
    - **Cours intensif sur l'analyse des données longitudinales**
      - **Les enquêtes longitudinales de Statistique Canada ;**
      - **Introduction à la gestion des données et à la création de variables au moyen du progiciel STATA ;**
      - **Description des données longitudinales ;**
      - **Éléments de modélisation ;**
      - **Analyse longitudinale (transitions, survie)**
      - **Plans complexes, pondération et variances robustes**