



General Social Survey
Enquête sociale générale

Vingt ans de données de l'Enquête sociale générale

Michael Wendt



Statistique
Canada

Statistics
Canada

Canada 



Aperçu

- Contexte
- Base de données historiques
- Modèle de données
- Exemples
- Conclusion



Statistique
Canada

Statistics
Canada

Canada 



Contexte

- Il existe 20 cycles de données complets
- Des guides de l'utilisateur, dictionnaires de données, descriptions de méthodologies, guides de pondération et d'estimation, etc., sont disponibles
- De nombreux articles de recherche ont été rédigés au fil des ans





Base de données historiques

- Une foule de données sont maintenant disponibles
 - 330 000 dossiers de répondants (quoique pour des années différentes), des milliers de variables sur chaque répondant
 - De plus, de l'information sur l'emploi du temps, les épisodes de victimisation, etc.
- L'idée est de regrouper les données en un format facile d'accès et d'offrir des outils pour les utiliser
- Offrir une base de données contenant des données sur 20 ans





Base de données historiques

- Pas une base de données regroupées mais des ensembles de données distincts, comprenant des variables harmonisées et des concepts présentés de manière uniforme
- Version ambitieuse de l'idée
 - « Regrouper » les fichiers des répondants par cycle
 - 330 000 rangées x des milliers de colonnes
 - Avoir un identificateur de l'année (= cycle = fichier)
 - Poids bootstrap
 - Offrir des liens vers d'autres fichiers existants, comme l'emploi du temps, les épisodes de victimisation





Base de données historiques

- Malheureusement, on ne peut pas simplement « empiler » les 20 ensembles de données les uns sur les autres
- Questions
 - Les variables peuvent avoir des noms, des formats, des ensembles de codes différents
 - Les méthodes de traitement ont changé au fil des ans (heureusement, pas trop)

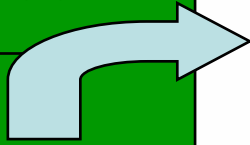
cycle	âge	sexe	...
1			
2			
3			
⋮			





Modèle de données

Cycle	Base
1	
2	
3	
4	
5	
...	



information sur les répondants commune à tous ou presque tous les cycles

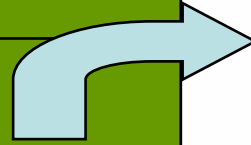
âge, sexe, état matrimonial, situation domestique du ménage, province, etc.





Modèle de données

	Base	Contrôle
Cycle		
1		
2		
3		
4		
5		
...		



identificateurs uniques,
métadonnées, variables
nouvellement établies,
contrôle méthodologique

numéro matricule,
pondération, variables
d'appariement, etc.





Modèle de données

	Base	Contrôle	Thèmes
Cycle			
1			
2			
3			
4			
5			
...			



des thèmes récurrents

temps consacré à l'aménagement de son terrain, etc.





Modèle de données

	Base	Contrôle	Thèmes	Spéciales
Cycle				
1				
2				
3				
4				
5				
...				

des sujets qui se retrouvent dans quelques cycles consécutifs seulement

utilisation d'Internet, etc.





Modèle de données

	Base	Contrôle	Thèmes	Spéciales	Autres
Cycle					Le reste ? ou !
1					
2					
3					
4					
5					
...					





Modèle de données : sommaire

	Base	Contrôle	Thèmes	Spéciales	Autres
Cycle					
1					
2					
3					
4					
5					
...					





Base de données historiques

- Permettra aux utilisateurs de :
 - suivre les estimations temporelles
 - créer une série chronologique d'estimations
 - suivre l'évolution des « cohortes » (personnes âgées de 25 ans en 1990, de 30 ans en 1995, etc.)
 - examiner des sous-populations semblables à différentes périodes (personnes âgées de 25 ans en 1990 et de 25 ans en 1995)
 - combiner de petits domaines
 - utiliser des cycles différents pour augmenter la taille de l'échantillon
 - en général, on ne peut regrouper toutes les données et, par exemple, examiner 160 000 femmes parce qu'il n'y a pas de véritable population de référence





Modèle de données : utilisation de variables

- De base
 - Estimations sur des longues séries chronologiques
 - Calibrage
 - Cohortes (un âge différent est employé pour une année différente) et domaines d'âge (année différente mais même âge)
- De contrôle
 - Créer la base de données
 - Vérifications de la qualité
 - Outils, comme l'estimation de la variance (bootstrap)
 - Appariement avec d'autres ensembles de données





Modèle de données : utilisation de variables

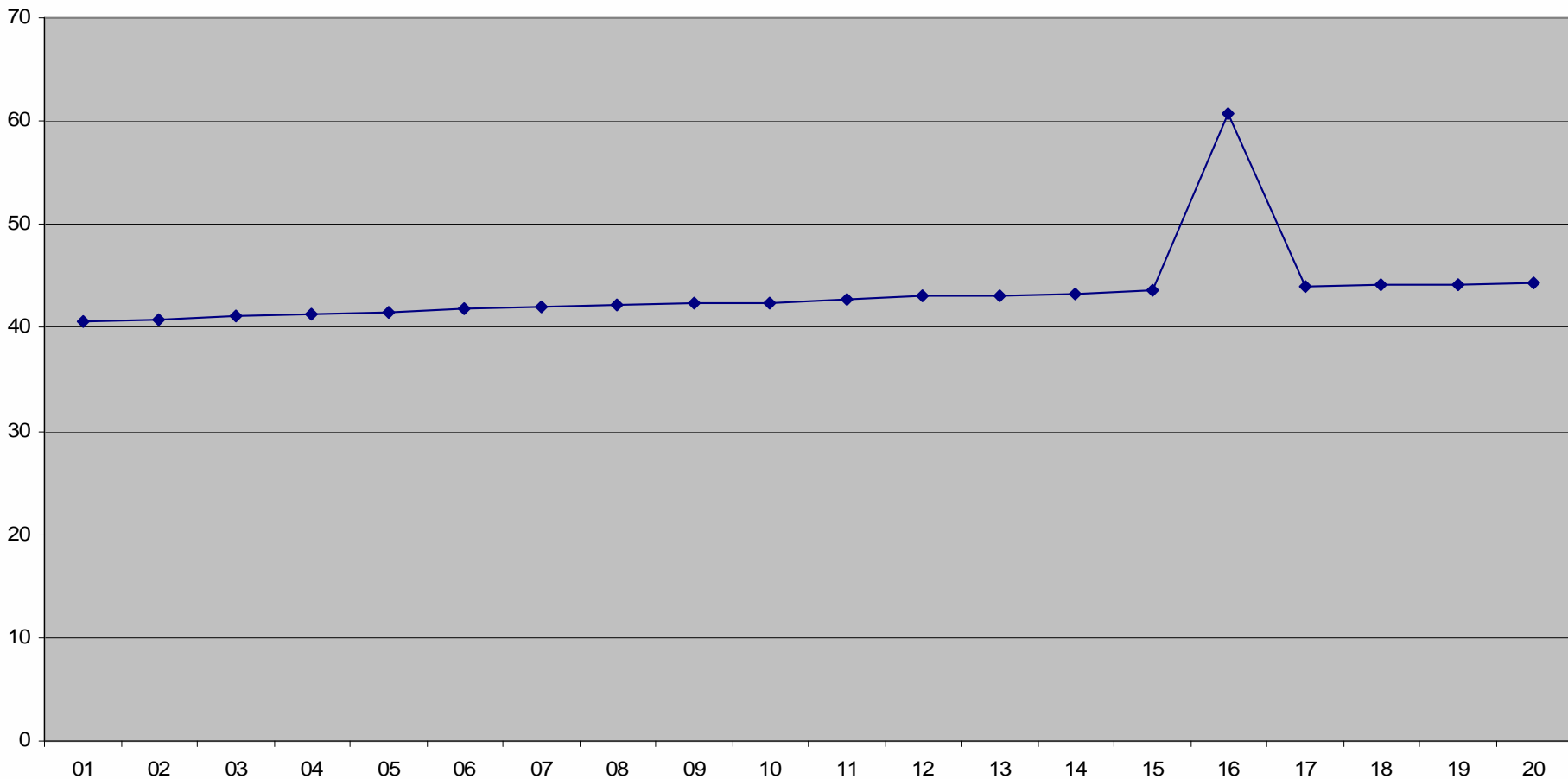
- Thèmes
 - Cohortes
 - Séries chronologiques avec moins de points mais plus de détails
- Spéciales
 - Petits domaines
 - Séries chronologiques très courtes
- Autres
 - Selon le type, la quantité, etc.





Exemple : estimations sous forme de série chronologique

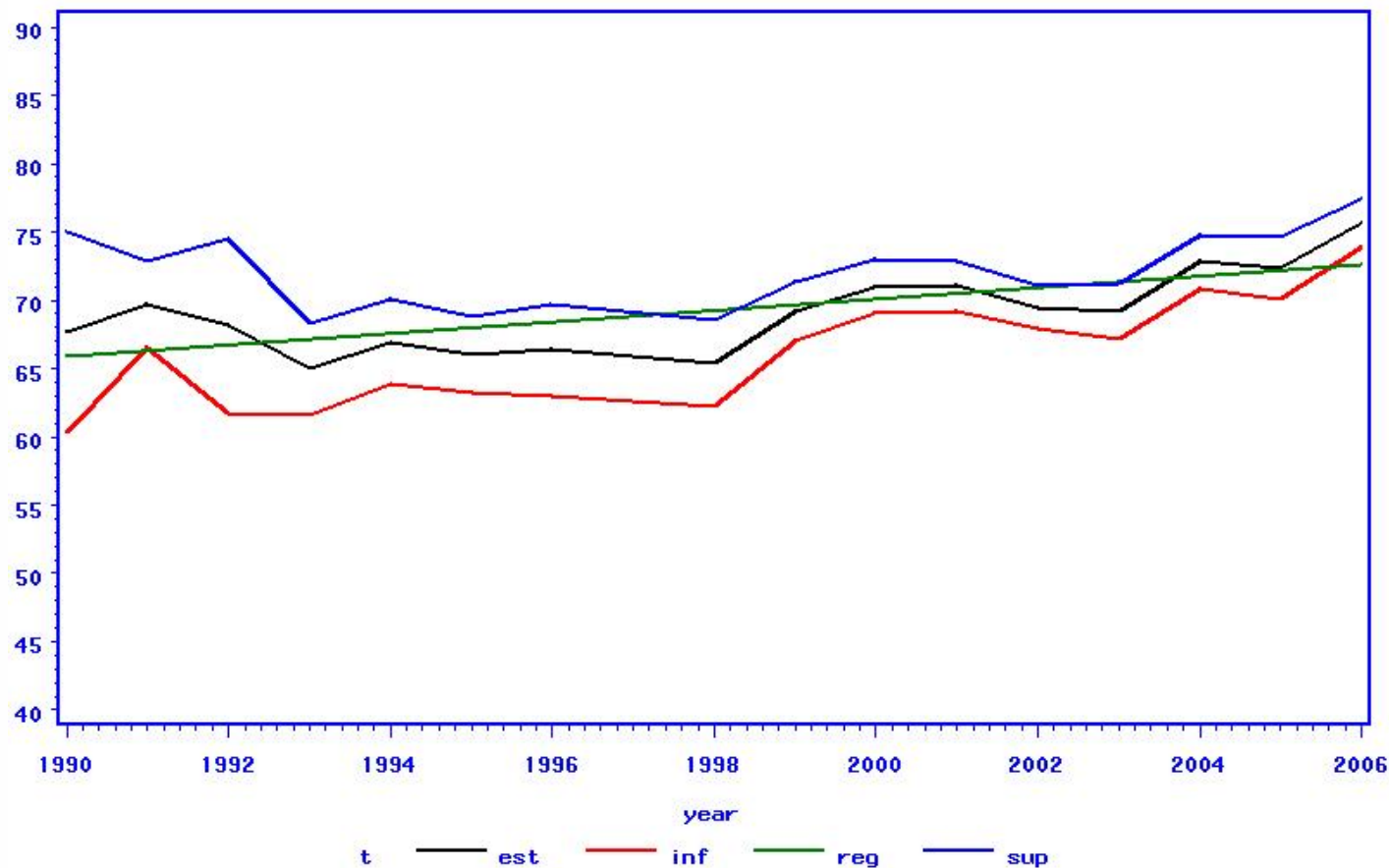
Âge moyen du répondant





Exemple : estimations sous forme de série chronologique

Pourcentage de dwellown = 1 pour 45+ au Quebec



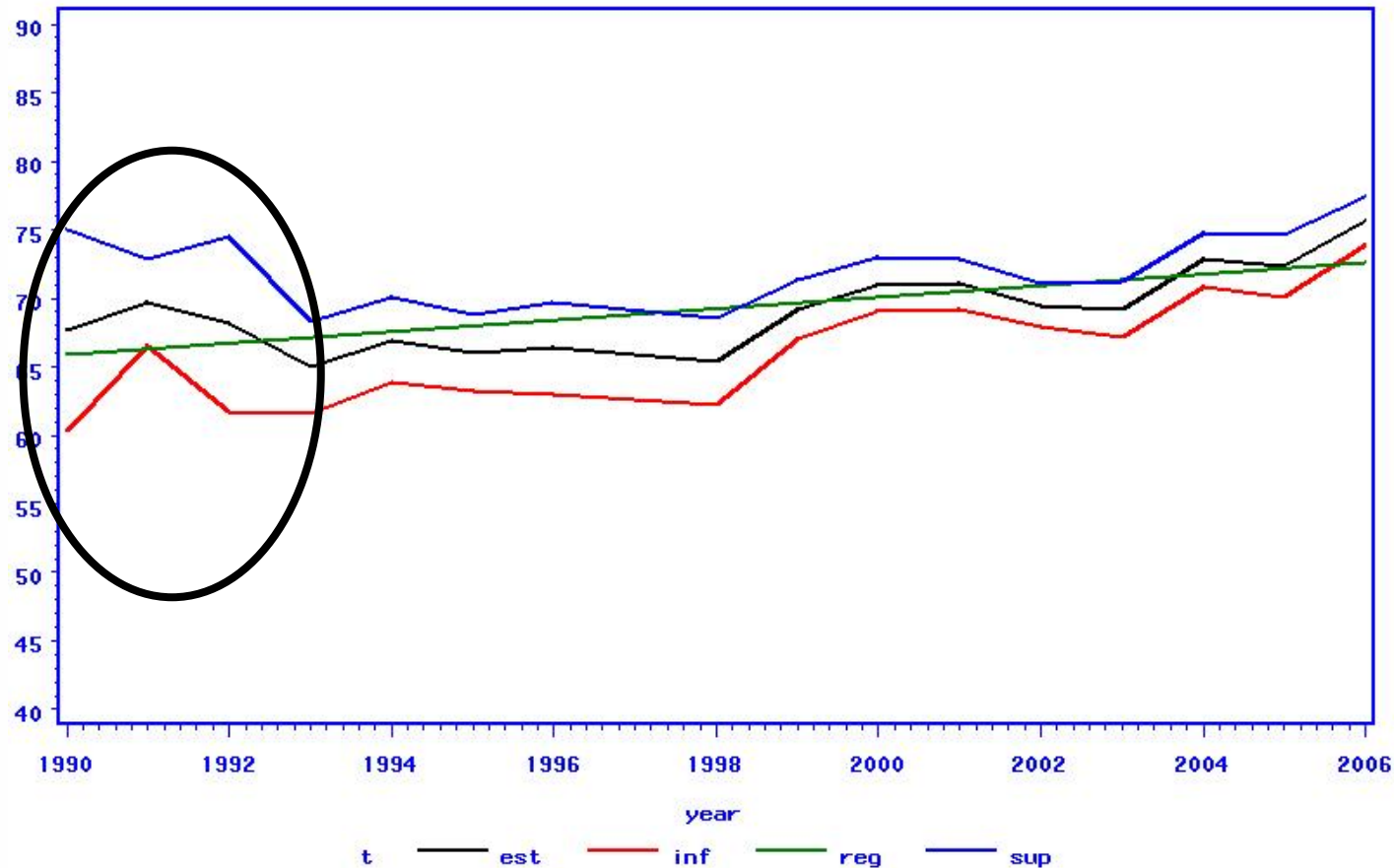
note: à des fins d'illustration seulement
source: Statistique Canada, Enquête sociale générale





Exemple : estimations sous forme de série chronologique

Pourcentage de dwellown = 1 pour 45+ au Quebec



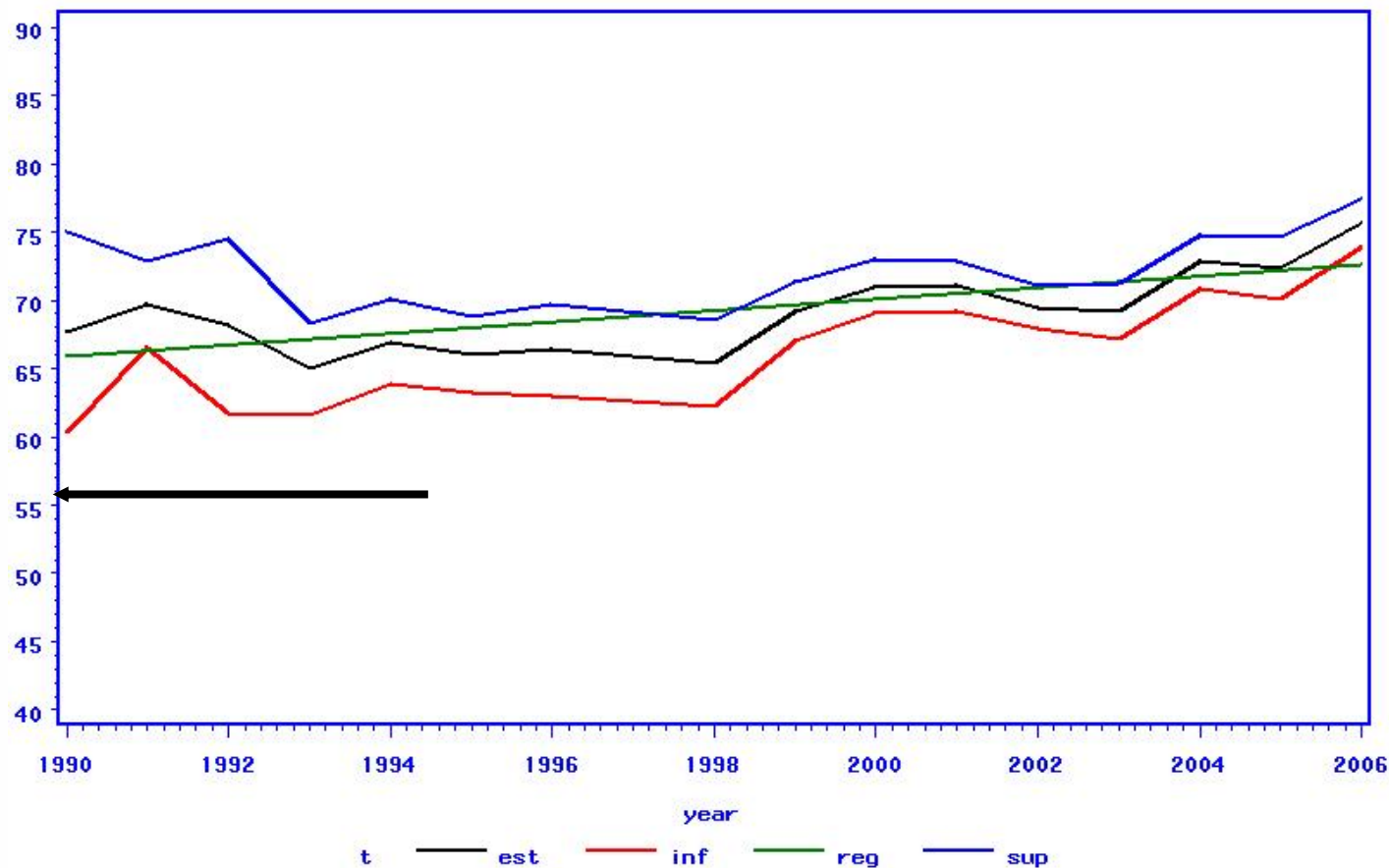
note: à des fins d'illustration seulement
source: Statistique Canada, Enquête sociale générale





Exemple : estimations sous forme de série chronologique

Pourcentage de dwellown = 1 pour 45+ au Quebec



note: à des fins d'illustration seulement
source: Statistique Canada, Enquête sociale générale





Exemple : combinaison de petits domaines

- Combiner des données de deux cycles ou plus
- Avantage : plus de données (problèmes de confidentialité, problèmes de variabilité)
- Nécessite des analyses non chronologiques ou quelques cycles peu chronologiques
- Il faut interpréter les estimations comme il se doit dans le contexte d'une population cible (selon le type d'estimation : paramètres descriptifs et du modèle)





Exemple : combinaison de petits domaines

- Personnes de 15 ans dont l'activité principale de la dernière année consistait à travailler à un emploi rémunéré, à leur propre compte ou à chercher un emploi rémunéré

Cycle	15 ans... etc.
13	moins de 10
14	moins de 10
15	moins de 10
total	plus de 10 oui!





Conclusion : produit

- Prévission : produit « nettoyé » dans les CDR d'ici le printemps 2008
- Mises à jour : ESG-21 et plus de variables
- Nous avons déjà une version « bêta »
 - Nous avons 896 variables
 - En voici des exemples : âge, cycle, propriétaire du logement, edu5, hsdsize, lanch, marstat, prov, recid_local, relig16, sex, wght_per
 - L'équipe est prête à créer des variables; des procédures ont été élaborées pour alimenter la base de données





Conclusion : produit

- Outils; travaux en cours
 - Poids bootstrap, ajout de variables d'intérêt pour les chercheurs, vérification de l'assurance de la qualité, etc.
- Documentation
 - Dictionnaire de données
 - Guides de l'utilisateur en format lisible
 - Renseignements sur les métadonnées
- On a commencé à faire de l'analyse (fiches d'information)





Conclusion : personne-ressource

- Questions?
- Personne-ressource :
 - Michael Wendt
 - 613-951-7314
 - michael.wendt@statcan.ca



Statistique
Canada

Statistics
Canada

Canada