

INSTITUT DE LA STATISTIQUE DU QUÉBEC
DIRECTION SANTÉ QUÉBEC

**Banque de données masquée
de l'Enquête sociale et de santé 1998**

Guide d'utilisation

Juillet 2001

INSTITUT DE LA STATISTIQUE DU QUÉBEC
DIRECTION SANTÉ QUÉBEC

**Banque de données masquée
de l'Enquête sociale et de santé 1998**

Guide d'utilisation

Nathalie Audet

TABLE DES MATIÈRES

Introduction	3
Avis concernant le masquage	4
Note sur la section IX du QRI.....	4
Contenu du fichier	5
Pondérations	6
Effets de plan	7

INTRODUCTION

L'enquête sociale et de santé 1998 a été réalisée par la Direction Santé Québec de l'Institut de la statistique du Québec. La population visée est l'ensemble des ménages privés de toutes les régions sociosanitaires du Québec, à l'exclusion des régions crie et inuite.

Le rapport général est publié sous le titre suivant :

Enquête sociale et de santé 1998, Québec, Institut de la statistique du Québec.

Le rapport est disponible sur le site WEB de l'Institut de la statistique du Québec :

<http://www.stat.gouv.qc.ca>

Ce guide s'adresse à l'utilisateur de la banque de données de l'Enquête sociale et de santé 1998.

La banque de données de distribution est non nominative et le contenu est légèrement masqué pour éviter des bris de confidentialité des répondants. Elle est distribuée à des fins de recherche et ne peut être utilisée à des fins commerciales.

DÉFINITIONS :

QAA : Questionnaire auto-administré destiné à tous les répondants de 15 ans et plus.

QRI : Questionnaire rempli par l'interviewer administré lors d'une entrevue en face à face auprès d'un répondant du ménage qui répond pour lui-même et pour chacun des membres du ménage.

INDICE : Créé à partir d'une ou plusieurs questions. Il peut être un simple regroupement de catégories ou construit selon les normes déjà existantes d'une échelle de mesure connue ou selon les instructions des chercheurs membres des groupes d'analyse de l'enquête.

LIVRE DE CODES : Cahier qui reprend les questionnaires de l'enquête en précisant les fréquences et proportions obtenues (non pondérées) pour chaque variable.

CAHIER TECHNIQUE, VOL 1 : Cahier qui explique la méthodologie et les aspects techniques de l'enquête.

CAHIER TECHNIQUE, VOL 2 : Cahier qui explique la construction des principaux indices.

AVIS CONCERNANT LE MASQUAGE

Un fichier de microdonnées dénominalisé et « légèrement masqué » a été produit pour les collaborateurs externes de l'ISQ afin qu'ils puissent bénéficier du potentiel analytique des données de l'Enquête sociale et de santé 1998. L'approche proposée par Willenborg et de Waal (1996) a été privilégiée car leur démarche rigoureuse est caractérisée par des règles statistiques du contrôle de la divulgation. Parmi les changements apportés, on retrouve la recatégorisation de variables, le regroupement de certaines valeurs extrêmes, la suppression locale de données, la suppression de variables trop discriminantes, ainsi que l'ajout d'un bruit aléatoire aux poids de sondage du QRI. Le cahier technique Vol. 2 présente les modifications des indices et le livre de codes présente les modifications des variables des questionnaires.

Les modifications apportées ont pour but de diminuer la possibilité qu'un répondant à l'enquête puisse être identifié, et par le fait même qu'on découvre de l'information sensible à son sujet. Afin de minimiser le contrôle statistique de la divulgation au niveau des individus, il a été décidé d'empêcher qu'un utilisateur puisse savoir à quel ménage appartient chaque individu du fichier. En conséquence, il n'est pas possible d'effectuer des analyses au niveau des ménages à l'aide du fichier de microdonnées « légèrement masqué ». Un des traitements particuliers appliqués aux microdonnées pour éviter la reconstitution des ménages est l'ajout d'un bruit aléatoire aux poids de sondage pour les variables du QRI puisque les poids originaux identifiaient les ménages; tous les autres poids du fichier n'ont subi aucun changement. Il est toutefois important de souligner qu'une analyse de l'impact sur les estimations de proportion des changements apportés aux variables (suppression locale aléatoire ou non et ajout de bruit aléatoire) a été effectuée. En effet, une vingtaine de variables provenant du QRI ou du fichier des indices ont été analysées, et des différences de proportion au niveau du Québec entre les estimations « avant changement » et « après changement » ont été détectées pour seulement 2 variables; de surcroît, ces différences étaient de l'ordre de 0,1 %. En conséquence, les perturbations apportées aux variables ont eu un impact minime sur les estimations au niveau du Québec.

WILLENBORG, L. et T. DE WAAL (1996), *Statistical Disclosure Control in Practice*, Springer.

NOTE SUR LA SECTION IX DU QRI

La section IX du QRI fournit des informations supplémentaires sur les problèmes de santé mentionnés lors de l'entrevue. Elle n'est pas incluse dans le fichier pour 2 raisons : elle ne doit pas être utilisée par individu tel que le fichier est constitué, il faut créer un fichier où chaque enregistrement présente un problème de santé plutôt qu'un individu. Autrement dit, l'unité d'analyse doit être le problème de santé, et une pondération spéciale doit être utilisée. Deuxièmement cette section alourdit considérablement le fichier de données par son ampleur. Cette section reste quand même disponible aux chercheurs qui en feront la demande.

CONTENU DU FICHIER:

QAA Nombre d'observations : 20 773 Nombre de variables : 348

QAA	Données du questionnaire QAA en format SAS
QAA_SPSS	Données du questionnaire QAA en format SPSS
DICT_QAA	Dictionnaire SAS pour le QAA, incluant formats et libellés des variables

QRI Nombre d'observations : 30 386 Nombre de variables : 404

QRI	Données du questionnaire QRI en format SAS
QRI_SPSS	Données du questionnaire QRI en format SPSS
DICT_QRI	Dictionnaire SAS pour le QRI, incluant formats et libellés des variables

INDICES (principaux indices et pondérations)

Nombre d'observations : 30 386 Nombre de variables : 196

INDICES	Indices en format SAS
IND_SPSS	Indices en format SPSS
DICT_INDICES	Formats SAS pour les indices.

La variable NOIDM pour identifier les enregistrements est commune à chaque fichier, elle permet de faire le lien entre les différents fichiers.

PROCÉDURE D'INSTALLATION POUR LA VERSION SPSS

Le fichier ENQ98_SPSS sur le CD de distribution contient tous les fichiers en format SPSS PORTABLE.

1. Créer un répertoire pour l'enquête sociale et de santé 1998, par exemple c:\ENQ98.
2. Sauvegarder le fichier ENQ98_SPSS dans c:\enq98
3. Dans SPSS, faire les opérations suivantes :

```
get sas / data='c:\enq98\enq98_spss' dset(qaa_spss).  
get sas / data='c:\enq98\enq98_spss' dset(qri_spss).  
get sas / data='c:\enq98\enq98_spss' dset(ind_spss).  
get sas / formats='c:\enq98\enq98_spss' fset(form).
```

Note : Les formats pour la version SPSS sont tous contenus dans le fichier FORM.

PONDÉRATIONS

Les pondérations doivent être intégrées à l'analyse pour faire de l'inférence à la population.

QRI

Les poids au QRI peuvent être utilisés dans les circonstances suivantes :

- Variables du QRI
- Croisements de variables du QRI

CPOIDS1P : Permet l'inférence à la population du Québec tant au niveau provincial¹ que par région². La valeur associée au répondant indique le nombre de personnes qu'il représente dans la population.

CPIDS1M1 : Pondération au QRI à utiliser si l'on veut effectuer des tests statistiques au niveau provincial¹. Les proportions obtenues par cette pondération procurent les mêmes résultats que la pondération POIDSQRI, mais apportent un changement à l'effectif observé. La somme des valeurs associées aux répondants est égale au nombre de répondants.

CPIM1R : Pondération au QRI à utiliser si l'on veut effectuer des tests statistiques par région². La somme des valeurs associées aux répondants est égale au nombre de répondants dans la région.

QAA

Les poids au QAA peuvent être utilisés dans les circonstances suivantes :

- Variables du QAA
- Croisements de variables du QAA
- Croisements entre variables du QAA et du QRI

POIDS2P : Permet l'inférence à la population de 15 ans et plus du Québec tant au niveau provincial¹ que par région². La valeur associée au répondant indique le nombre de personnes qu'il représente dans la population.

POIDS2M1 : Pondération au QAA à utiliser si l'on veut effectuer des tests statistiques au niveau provincial¹. Les proportions obtenues par cette pondération procurent les mêmes résultats que la pondération POIDSQAA mais apportent un changement à l'effectif observé. La somme des valeurs associées aux répondants est égale au nombre de répondants au QAA.

P2M1R : Pondération au QAA à utiliser si l'on veut effectuer des tests statistiques par région. La somme des valeurs associées aux répondants est égale au nombre de répondants au QAA dans la région.

¹Au niveau provincial : Lorsque les analyses concernent l'ensemble de la population du Québec.

²Par région : Lorsque les analyses concernent une seule région.

EFFETS DE PLAN

Étant donné le plan d'échantillonnage complexe de l'enquête, il n'est pas suffisant d'utiliser les poids ramenés à la taille de l'échantillon lorsque des tests statistiques sont effectués : on doit en plus apporter un ajustement à l'aide d'un effet de plan.

Les effets de plan doivent être intégrés dans les calculs du khi deux et dans l'estimation de la précision de la proportion. Les précisions et les tests ainsi obtenus sont approximatifs puisqu'ils utilisent un effet de plan moyen obtenu pour un ensemble de statistiques au lieu de l'effet de plan associé à la statistique désirée. Pour plus de détails, voir le cahier méthodologique et technique, vol 2. Ils n'ont pas été vérifiés pour d'autres tests statistiques ou d'autres estimations, il est possible qu'ils ne soient pas adéquats dans ces circonstances.

La meilleure façon d'intégrer l'effet de plan est de diviser la pondération par l'effet de plan avant toute opération. Ensuite, on utilise la valeur résultante comme pondération dans les analyses.

Ex. Supposons le croisement de 2 variables du QAA et que le test de khi deux est désiré. La pondération à moyenne 1 doit être divisée par l'effet de plan avant toute opération.

Pondération à moyenne 1 pour le QAA : POIDS2M1

Effet de plan général : 2

Pondération à utiliser dans la fréquence : POIDS2M1/2

L'effet de plan « Général » est utilisé pour toutes les variables et tous les indices du questionnaire sauf pour les exceptions mentionnées dans les tableaux suivants.

Effets de plan pour les analyses provinciales

QRI	Effet de plan
Général	2,1
Sections 12, 13 et 15 (sauf questions 175 à 178) SECALIM, CAUS, COUVMED, COUVPRI, COUVSDEN, COUVSH, COUVVUE, TYPEPEG Question 157, SUFREV98	5,8
Questions 175 à 178	8,8
Questions 41, 48 et 75, MATCONJ2, VACCIN, VACREC	3,3

QAA	Effet de plan
Général	2
Questions 223 à 225, 228 et 229, GROUPTH2, LANGMAT	4,8
Questions 9 (a, b et c), 157 à 159, 161, 171 à 174 194, 195 et 232 à 238 REPASEXA, REPASEXT, VSPIRIT, SSPIRIT, RELENFA, ISSAV, ISUTIL, DURPAUV	2,6

Effets de plan pour les analyses régionales - QRI

QRI	Effet de plan par Région															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Général	1,5	1,4	1,4	1,7	1,4	1,5	1,7	1,6	1,3	1,1	1,7	1,2	1,5	1,5	1,5	1,8

QRI	Effet de plan par Région			
	1, 2, 3, 5, 6, 8, 9, 12, 13, 14, 15	4, 11	7, 16	10
Questions 41, 48, 75b et c MATCONJ2	1,8	1,7	2,2	1,4
Section 12, 13 et 15 (sauf questions 175 à 178) SECALIM, CAUS, COUVMED, COUVPRI, COUVSDEN, COUVSH, COUVVUE, TYPereg SUFREV98, FAMPAR	4,4	4,4	4,4	4,4
Questions 175 à 178	6,6	6,6	6,6	6,6

Effets de plan pour les analyses régionales – QAA

QAA	Effet de plan par Région															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Général	1,5	1,3	1,4	1,5	1,4	1,4	1,7	1,2	1,4	1,2	1,6	1,3	1,4	1,4	1,4	1,7

QAA	Effet de plan par Région			
	1, 2, 3, 4, 5, 6, 8, 9, 12, 13, 14, 15, 20	7	10	11
DURPAUV, PERCU Questions 9a, b et c, 157, 158, 194, 195, 229, 238 REPASEXA, REPASEXT, RELACT, VSPIRIT, SSPIRIT, ISSAV, ISUTIL	1,8	2,1	1,4	2

GROUPTH2 : Régions 6, 7, 13 et 16 seulement : Effet de plan de 3,9