

Les implications de l'arrondissement aléatoire dans l'exploitation des données du recensement : quelques exemples

Yolande Lavoie et Pierre Levasseur

Volume 4, numéro 1, 1975

URI : <https://id.erudit.org/iderudit/305793ar>

DOI : <https://doi.org/10.7202/305793ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Association des démographes du Québec

ISSN

0380-1721 (imprimé)

1705-1495 (numérique)

[Découvrir la revue](#)

Citer cet article

Lavoie, Y. & Levasseur, P. (1975). Les implications de l'arrondissement aléatoire dans l'exploitation des données du recensement : quelques exemples. *Cahiers québécois de démographie*, 4(1), 10–27. <https://doi.org/10.7202/305793ar>

Les implications de l'arrondissement aléatoire dans l'exploitation des données du recensement : quelques exemples.

PAR

Yolande LAVOIE* et Pierre LEVASSEUR**

Nous sommes tributaires de Statistique Canada qui collecte d'une part l'information fournissant les populations de départ pour les perspectives et d'autre part, les données sociales les plus couramment utilisées dans les études que nous avons à produire.

Or cet organisme met à la disposition des "utilisateurs" du recensement canadien de 1971 des tableaux sommaires sur bande d'ordinateur, afin qu'ils puissent en tirer les analyses de leur choix. C'est ainsi que les services de diffusion provinciaux (1) ont pu se procurer les bandes sommaires AI DEM 001 et AI DEM 002 présentant, sous une forme abrégée (001) ou sous une forme détaillée (002), les caractéristiques démographiques de la population canadienne suivant les secteurs de dénombrement. Ces secteurs sont ainsi définis: "Unité géographique recensée par un représentant du recensement et pouvant comprendre jusqu'à 300 ménages, mais toujours moins de 100 fermes, c'est-à-dire une population d'environ 500 personnes".

(*) Service de la démographie et du recensement B.S.Q.-Québec-

(**) Service de la Recherche, ministère des Affaires Sociales -Québec-

(1) Au Québec: Bureau de la Statistique du Québec, division de la diffusion des données du recensement

Peu d'utilisateurs, qu'ils soient chercheurs, planificateurs ou technocrates, font directement usage de données présentées suivant ce découpage géographique très fin. En général, les utilisateurs agrégeront les secteurs de dénombrement pour reconstituer les unités géographiques qui conviennent à leurs besoins.

Les bandes sommaires et l'arrondissement aléatoire

Selon la technique de l'arrondissement aléatoire, tous les tableaux mis à la disposition du public sont arrondis à la base 5 et le sens vers lequel on arrondit est déterminé par le hasard plutôt que par une série explicite de règles. Le chiffre 5 d'un tableau indique la valeur réelle se situe entre 1 et 9.⁽¹⁾

Imaginons maintenant la façon dont se présentera un tableau répartissant 500 personnes environ suivant le sexe puis, à l'intérieur de cette première division, suivant l'âge (année d'âge). Il est évident que peu de cellules contiendront des nombres supérieurs à 10 puisqu'une telle distribution répartit les 500 personnes dans quelque 200 cellules. Qu'on se représente un tel tableau après qu'il aura été soumis à l'arrondissement aléatoire dont "l'influence sur chaque cellule de tableau est inversement proportionnelle à la taille de cette cellule."⁽²⁾ Il est évident que n'y apparaîtront que les 3 ou 4 premiers

(1) "L'arrondissement aléatoire: Technique de protection contre la divulgation illégale de renseignements dans les tableaux de recensement", Statistique Canada, Division du recensement, note de recherche préparée par D.M. Murphy, 2 août 1972, 11 p.

(2) "L'arrondissement aléatoire: ...", p. 11

multiples de "5" et des "0".

Un tel tableau ne permet de calculer ni une distribution relative suivant le sexe et l'âge (qu'on songe à l'allure de la pyramide des âges qui en découlerait), ni des rapports de masculinité pour ne citer que les calculs les plus simples et les plus courants.

Or, si Statistique Canada met les bandes sommaires à la disposition du public par le truchement des services de diffusion provinciaux, c'est qu'il juge que ces bandes constituent une banque d'information plus détaillée et plus raffinée que le contenu des brochures du recensement. Il arrive en effet que la reconstitution de certaines unités géographiques ne soit possible qu'à partir des secteurs de dénombrement qui constituent le niveau le plus fin de découpage territorial.

EXEMPLE 1: Un comté, les Iles-de-la Madeleine

Voyons ce qui se passera si, à partir de la bande A1 DEM 002 (secteurs de dénombrement et année d'âge), on veut reconstituer la population d'un comté, en l'occurrence les Iles-de-la-Madeleine, et la représenter sous la forme d'un tableau distribuant les effectifs suivant le sexe et le groupe quinquennal d'âges. Rappelons que l'arrondissement a été fait au niveau des âges individuels, par sexe, pour chacun des secteurs de dénombrement.

Le tableau 1 présente d'une part, les données provenant de la bande Al DEM 002 et d'autre part, celles qui sont contenues dans la publication de recensement "Population. Groupes d'âge. cat. 92-715, vol. 1 - part: 2" où l'arrondissement a été fait au niveau du comté, par sexe et groupe quinquennal d'âges. Quant au tableau 2, il contient les rapports de masculinité correspondant à chacune des distributions du tableau 1.

Tableau 1: Population suivant le sexe et le groupe d'âges quinquennal. Iles-de-la-Madeleine. 1971.

Groupes d'âges	bande A1 DEM 002				publication 92-715			
	nombres absolus		pour 100 de population totale		nombre absolus		pour 100 de population totale	
	H	F	H	F	H	F	H	F
0-4	680	715	5.1	5.4	710	665	5.3	5.0
5-9	935	870	7.0	6.5	905	875	6.8	6.6
10-14	910	865	6.8	6.5	895	890	6.7	6.7
15-19	805	815	6.0	6.1	840	805	6.3	6.0
20-24	605	535	4.5	4.0	645	515	4.8	3.9
25-29	480	395	3.6	3.0	445	425	3.3	3.2
30-34	380	345	2.8	2.6	385	350	2.9	2.6
35-39	315	375	2.4	2.8	355	320	2.7	2.4
40-44	310	270	2.3	2.0	300	275	2.3	2.1
45-49	275	170	2.1	1.3	260	225	2.0	1.7
50-54	250	205	1.9	1.5	255	220	1.9	1.7
55-59	255	285	1.9	2.1	240	235	1.8	1.8
60-64	215	195	1.6	1.5	180	190	1.3	1.4
65-69	180	125	1.3	0.9	170	150	1.3	1.1
70-74	105	110	0.8	0.8	115	120	0.9	0.9
75-79	60	100	0.5	0.8	75	85	0.6	0.6
80-84	35	80	0.3	0.6	60	35	0.4	0.3
85-89	35	20	0.3	0.2	30	35	0.2	0.3
90-94	15	10	0.1	0.1	15	15	0.1	0.1
95-99	5	-	-	-	5	5	-	-
100+	-	-	-	-	-	-	-	-
Σ des colonnes	6 850	6 485	51.4	48.6	6 885	6 435	51.7	48.3
Grand total "officiel"	6 865	6 440			6 870	6 435		

Tableau 2. Rapports de masculinité. Iles-de-la-Madeleine, 1971.

Groupes d'âges	bande A1 DEM 002	publication 92-715
0-4	95.1	106.8
5-9	107.5	103.4
10-14	105.2	100.6
15-19	98.8	104.3
20-24	113.1	125.2
25-29	121.5	104.7
30-34	110.1	110.0
35-39	84.0	110.9
40-44	114.8	109.1
45-49	161.8	115.6
50-54	122.0	115.9
55-59	89.5	102.1
60-64	110.3	94.7
65-69	144.0	113.3
70-74	95.5	95.8
75-79	60.0	88.2
80-84	43.8	171.4
85-89	175.0	85.7
90-94	150.0	100.0
95-99	-	100.0
100+	-	
Ensemble	106.6	106.8
0-14	103.1	103.3
15-34	108.6	110.5
35-64	108.0	108.5
65+	97.8	105.6

D'énormes distorsions apparaissent dans la distribution qui résulte de l'agrégation des secteurs de dénombrement de la bande A1 DEM 002. La pyramide des âges dressée à partir des données du tableau 1 manifeste bien ces distorsions de la structure par âge et par sexe (figure 1). De même, la figure 2 indique que la répartition suivant le sexe est faussée et que l'interprétation qu'on en aurait pu faire concernant la migration différentielle selon le sexe ou concernant les marchés matrimoniaux serait loin de traduire la réalité.

L'importance des écarts entre la série A1 DEM 002 et la série 92-715 figure au tableau 3.

Tableau 3: Ecart en nombres absolus et en nombres relatifs entre les résultats de la bande A1 DEM 002 et ceux de la brochure 92-715. Population distribuée par groupe d'âges et par sexe. Iles-de-la-Madeleine, 1971

Groupe d'âges	Nombres absolus		Nombres relatifs (%)	
	H	F	H	F
0-4	-30	+50	- 4	+ 3
5-9	+30	- 5	+ 3	- 1
10-14	+15	-25	+ 2	- 3
15-19	-35	+10	- 4	+ 1
20-24	-40	+20	- 6	+ 4
25-29	+35	-30	+ 3	- 7
30-34	- 5	- 5	- 1	- 1
35-39	-40	+55	-11	+17
40-44	+10	- 5	+ 3	- 2
45-49	+15	-55	+ 6	-24
50-54	- 5	-15	- 2	- 7
55-59	+15	+50	+ 6	+21
60-64	+35	+ 5	+19	+ 3
65-69	+10	-25	+ 6	-17
70-74	-10	-10	- 9	- 8
75-79	-15	+15	-20	+18
80-84	-25	+45	-42	+129
85-89	+ 5	-15	+17	-43
90-94	0	- 5	0	-33

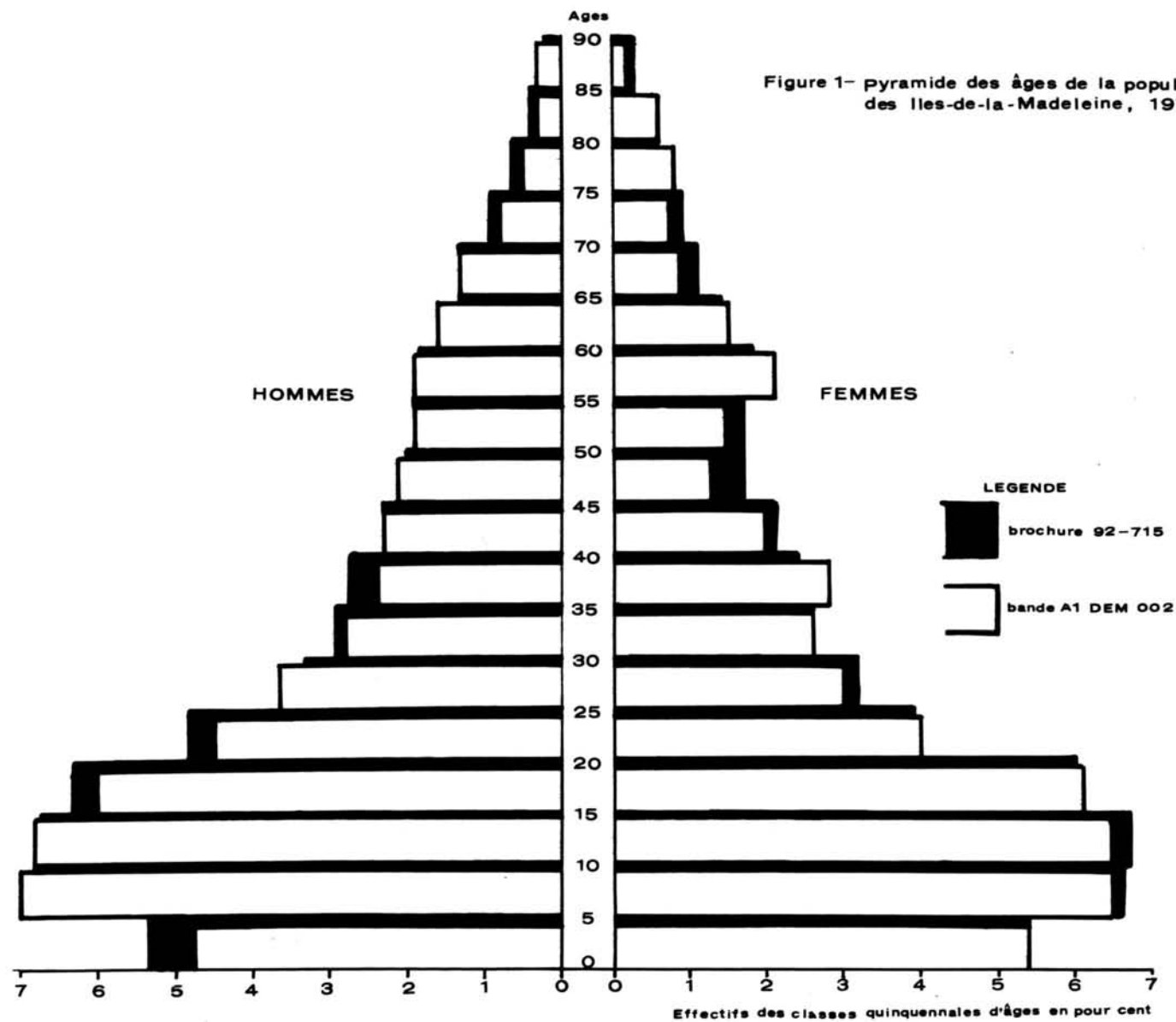
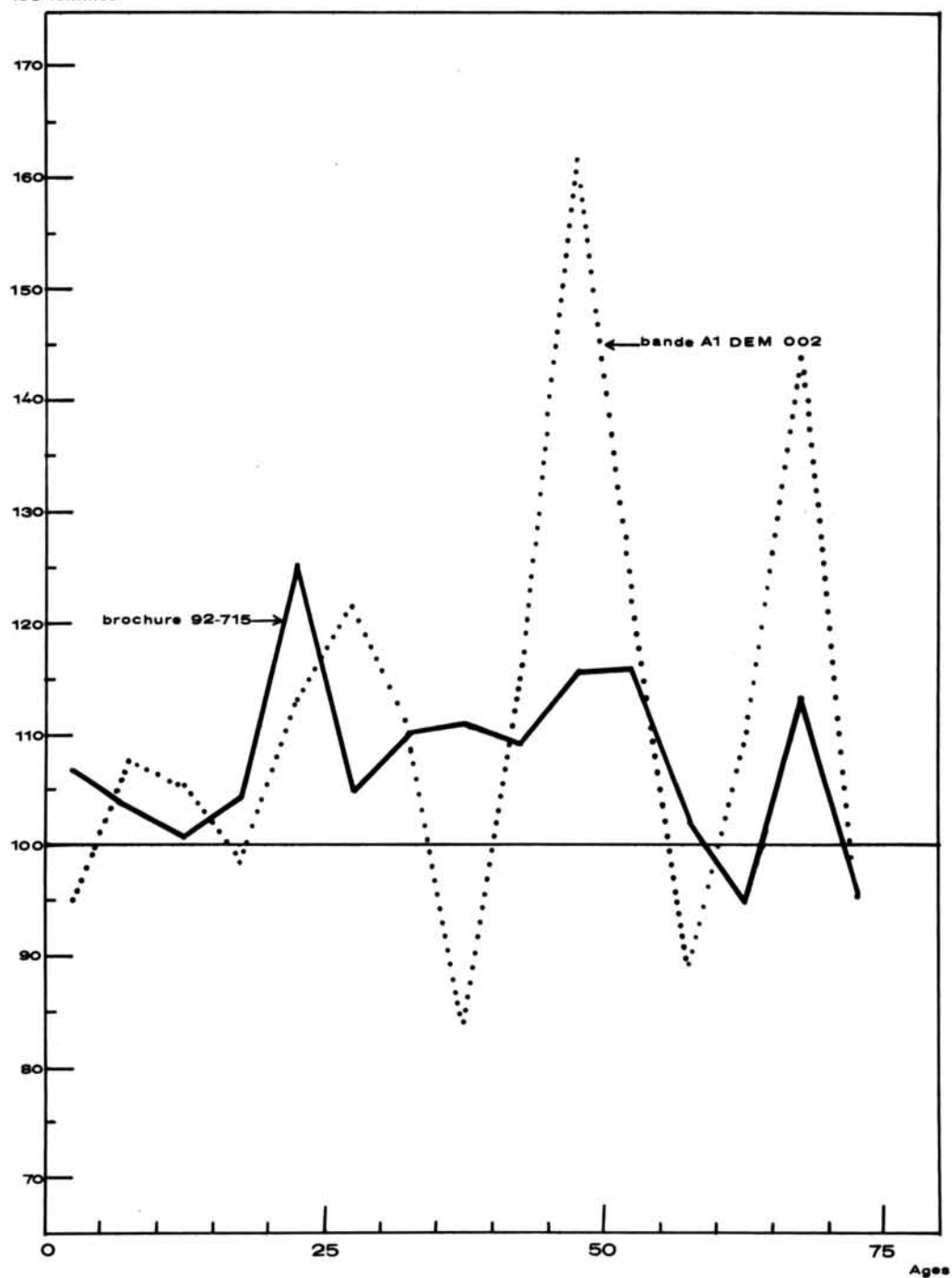


Figure 2- Rapports de masculinité, Iles-de-la-Madeleine,
- 1971 -

Hommes pour
100 femmes



Le sens des écarts, tout comme celui de l'arrondissement, est aléatoire. Dans plus de la moitié des cas, les écarts entre les deux séries sont supérieurs à 5%. Si on excepte les 70 ans et plus, dont les effectifs sont très faibles, on constate que les différences entre la bande et la brochure dépassent dans quelques cas 20%. Il est évident que l'analyse des données provenant de l'agrégation des secteurs de dénombrement, préalablement soumis à l'arrondissement aléatoire, constitue une entreprise pleine d'embûches. En effet, si dans notre exemple et c'est volontaire, nous disposons d'une série de données justes, il n'en serait pas de même si nous avions reconstitué non plus un comté, mais un bassin hydrographique, une région administrative, un district judiciaire ou toute autre unité territoriale pour laquelle le recensement ne publie pas de statistiques. Nous serions alors privés d'un terme de comparaison permettant de juger de la validité des données.

Il ressort de ce test sur le comté des Iles-de-la-Madeleine, que les bandes sommaires sont d'une utilité douteuse quant on doit reconstituer des territoires en agrégeant des secteurs de dénombrement. L'exemple que nous avons choisi n'a rien d'un cas exceptionnel. Cependant, du fait de la population restreinte de ce comté, les écarts entre la série A1 DEM 002 et la brochure 92-715 prennent une importance relative considérable. Il ne faut toutefois pas oublier que nous avons souvent à reconstituer des ensembles de cette taille.

EXEMPLE 2: Les régions administratives ou socio-sanitaires.

La plupart des démographes oeuvrant au Québec, ont eu ou auront à travailler avec les données relatives à la population des régions administratives ou socio-sanitaires distribuée par âge et par sexe. Mais les séries varient selon qu'elles proviennent des bandes sommaires, A1 DEM 001 et 002 (B.S.Q.) ou de commandes spéciales à Statistique Canada. Dans ce dernier cas, l'arrondissement aléatoire garantissant la confidentialité est effectué seulement au niveau de la région administrative ou socio-sanitaire.

Nous avons comparé les données (sexe masculin) de nos deux sources pour quatre régions socio-sanitaires:

- A) Bas St-Laurent/Gaspésie (niveau de 100,000 h).
- B) Québec (niveau de 500,000 h).
- C) Montréal Métropolitain (niveau de 1,000,000 h).
- D) Laurentide Lanaudière (niveau de 200,000 h).

TABLEAU 4

TABLEAU COMPARATIF¹ DE LA DISTRIBUTION SELON L'ÂGE DE LA POPULATION DE
CERTAINES RÉGIONS DU QUÉBEC (JUIN 1971) SUIVANT DEUX NIVEAUX D'ARRONDISSEMENT

SEXE MASCULIN

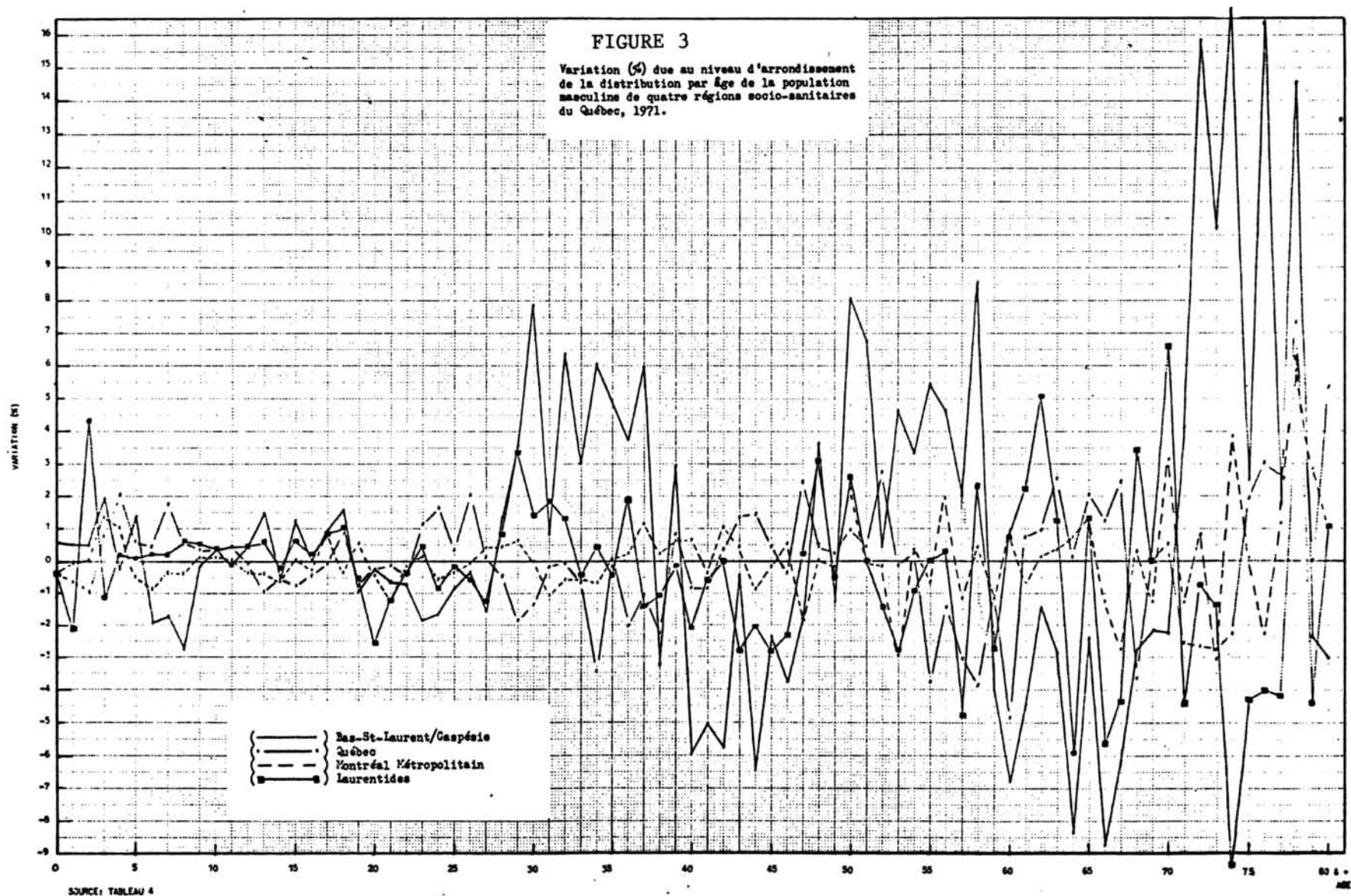
(VARIATION %)

AGE	BAS-ST-LAURENT	QUEBEC	MONTREAL METROPOLITAIN	LAURENTIDES	AGE	BAS-ST-LAURENT	QUEBEC	MONTREAL METROPOLITAIN	LAURENTIDES
0	0.560	-0.342	-0.381	-0.341	42	-5.702	0.395	1.042	0.000
1	0.501	-0.602	-0.063	-2.091	43	-0.418	1.318	0.462	-2.820
2	0.500	-0.997	0.032	4.314	44	-6.452	1.401	-0.840	-2.069
3	1.909	0.786	1.350	-1.189	45	-2.326	0.490	-0.110	-2.878
4	-0.216	2.006	1.040	0.266	46	-3.797	-0.302	0.519	-2.370
5	1.370	0.605	-0.437	0.119	47	-1.826	2.451	-1.812	0.246
6	-1.920	0.492	-0.870	0.215	48	3.604	0.408	0.078	3.093
7	-1.727	1.796	-0.372	0.213	49	-1.266	0.203	-0.191	-0.503
8	-2.754	0.558	-0.311	0.607	50	8.021	0.999	2.275	2.611
9	-0.161	0.328	0.189	0.522	51	6.731	0.448	-0.043	0.000
10	0.316	0.277	0.142	0.392	52	0.481	2.768	-0.199	-1.471
11	0.472	0.462	0.095	-0.100	53	4.615	-0.115	-2.960	-2.769
12	0.467	-0.093	-0.311	0.497	54	3.333	0.242	0.350	-0.965
13	1.495	-0.910	-0.360	0.615	55	5.405	-3.752	-0.608	0.000
14	-0.597	-0.593	-0.639	-0.208	56	4.639	-1.410	1.987	0.327
15	1.236	-0.784	0.099	0.666	57	2.062	-3.063	-1.143	-4.792
16	0.000	-0.412	-0.244	0.222	58	8.571	-3.815	0.472	2.310
17	0.940	-0.096	0.890	0.838	59	-4.000	-1.128	-1.061	-2.787
18	1.596	0.857	-0.313	1.013	60	-6.780	-4.864	0.820	0.719
19	-0.958	-0.791	0.575	-0.552	61	-4.430	0.770	-0.667	2.290
20	-0.217	-0.269	-0.432	-2.564	62	-1.439	0.984	0.126	5.099
21	-0.690	-0.111	-1.238	-1.264	63	-2.899	2.551	0.334	1.250
22	-0.733	-0.406	-0.126	-0.323	64	-8.333	0.177	0.703	-5.932
23	-1.842	1.137	0.146	0.491	65	-2.439	2.079	1.396	1.357
24	-1.676	1.614	-0.523	-0.825	66	-8.730	1.224	-1.240	-5.603
25	-0.882	0.322	-0.310	-0.166	67	-6.140	2.412	-2.752	-4.327
26	-0.318	2.067	-0.027	-0.655	68	-2.778	-3.687	0.286	3.448
27	-1.567	0.133	0.413	-1.260	69	-2.128	0.000	-1.259	0.000
28	1.329	-0.499	0.432	0.817	70	-2.222	0.548	3.175	6.579
29	3.534	-1.883	0.611	3.398	71	4.167	-2.571	-1.264	-4.375
30	7.865	-1.365	0.000	1.421	72	15.854	-2.736	0.840	-0.709
31	0.837	-0.165	-1.053	1.890	73	10.256	-2.694	-3.030	-1.399
32	6.329	-0.088	-0.599	1.365	74	16.883	-2.273	3.871	-9.402
33	3.030	-0.657	-0.598	-0.408	75	2.500	1.961	0.181	-4.237
34	6.047	-3.441	-0.638	0.428	76	16.393	3.084	-2.222	-4.000
35	4.911	0.096	-0.108	-0.409	77	1.786	2.577	1.129	-4.124
36	3.756	-2.069	0.290	1.911	78	14.583	7.303	5.897	6.250
37	5.963	-1.082	1.179	-1.471	79	-2.326	0.000	2.849	-4.478
38	-3.239	-2.218	0.245	-1.027	80+	-3.020	5.316	1.026	1.050
39	2.917	0.824	0.648	-0.196					
40	-5.963	-0.878	0.699	-2.066					
41	-5.063	-0.856	-0.314	-0.636					
					TOTAL	0.348	0.016	0.002	-0.221

(1) Résultats portés sur ce tableau = $\left(\frac{A - B}{A} \right) \times 100$

où

A= Population, distribuée par âge et par sexe, obtenue de Statistique Canada (arrondis faits, au niveau de la région).
B= Population, distribuée par âge et par sexe, obtenue du B.S.Q. (arrondis faits, au niveau du secteur de dénombrement).



Les résultats de cette comparaison ont été portés au tableau 4, et sont illustrés par la figure 3. Ils nous suggèrent les quelques remarques qui suivent:

1. Au niveau des populations totales, les écarts sont minimes.
2. La valeur relative des écarts varie rapidement d'un âge à l'autre, peu importe la région. Il n'y a donc aucune correction systématique possible.
3. Plus la population de la région est faible, plus les différences sont importantes au niveau de la distribution par âge, la cause en étant sans doute que le nombre restreint de secteurs de dénombrement, dans une petite région, ne permet pas l'annulation des écarts dus aux arrondis.
4. C'est surtout aux âges élevés que les écarts prennent des proportions inquiétantes, avec toutes les conséquences qui pourraient en découler au niveau des politiques d'aide aux personnes âgées.

EXEMPLE 3: L'ensemble du Québec

Même au niveau de l'ensemble du Québec, les distorsions entre les deux séries sont gênantes. Les écarts, à certains âges, ne sont pas négligeables (voir tableau 5). Le total des groupes d'âge s'écarte

Tableau 5: Ecartis en nombres absolus et en nombres relatifs entre les résultats de la bande A1 DEM 002 et ceux de la brochure 92-715. Ensemble du Québec, 1971.

Groupes d'âges	nombres absolus		pour cent	
	H	F	H	F
0-4	- 310	+ 15	- 0.13	+ 0.01
5-9	+ 80	+ 250	+ 0.02	+ 0.08
10-14	+ 835	+ 910	+ 0.24	+ 0.28
15-19	- 1030	+ 195	- 0.33	- 0.06
20-24	20	- 305	- 0.01	+ 0.11
25-29	45	- 105	- 0.02	+ 0.04
30-34	+ 270	+ 500	+ 0.14	+ 0.27
35-39	- 635	+ 540	- 0.35	+ 0.30
40-44	- 540	- 110	- 0.30	- 0.06
45-49	+ 795	- 320	+ 0.48	- 0.19
50-54	+ 305	+ 65	+ 0.22	+ 0.04
55-59	+ 360	+ 745	+ 0.29	+ 0.57
60-64	- 125	+ 25	- 0.13	+ 0.02
65-69	+ 115	+ 50	+ 0.16	+ 0.06
70-74	+ 485	- 630	+ 0.98	- 0.99
75-79	+ 355	+ 585	+ 1.14	+ 1.38
80-84	- 495	+ 65	- 2.92	+ 0.26
85-89	+ 15	+ 50	+ 0.21	+ 0.45
90-94	-	- 50	-	- 1.57
95+	-	-	-	-
Total "officiel"	- 300	+ 15	- 0.01	-
Σ des groupes d'âges	+ 415	+1905	+ 0.01	+ 0.06

sensiblement de celui qui provient de la sommation des totaux des secteurs de dénombrement (+1,905 pour le sexe féminin); le calcul des distributions relatives s'en trouve faussé. De plus, on possède pour le Québec autant de données de population totale et de répartitions de la population qu'il y a de niveaux d'arrondissement. Il s'ensuit qu'on doit multiplier les mises en garde, les notes explicatives, les estimations des marges d'erreur dans toutes les études faites à partir des données préalablement arrondies.

Nous nous en tenons évidemment ici aux seuls problèmes créés par l'arrondissement aléatoire et nous ne les discutons qu'à un niveau assez grossier puisque nous ne considérons toujours que les variables sexe et âge. Il convient toutefois de souligner que s'y ajoutent, pour les variables saisies par échantillonnage, les incohérences créées par les procédés de gonflement de l'échantillon pour reconstituer la population totale. Dans ces cas, à moins de disposer de données de références correctes, il est impossible de faire la part de ce qui revient comme erreur à l'arrondissement aléatoire d'une part et au passage de l'échantillon à la population totale d'autre part.

CONCLUSION.

Les citoyens canadiens peuvent dormir tranquilles, même les astuces des plus rusés d'entre les utilisateurs des données du recensement ne sauraient arracher aux tableaux provenant du recensement de 1971 le moindre indice permettant d'identifier un individu. Par contre, Statistique Canada s'expose à une avalanche de demandes spéciales dont le résultat est d'étendre les délais entre la formulation de la demande et la réception des résultats. Dans un deuxième temps, on assistera peut-être à un désintéressement croissant des utilisateurs qui se seront résignés à user d'expédients pour arriver à satisfaire leurs besoins.

Cette expérience avec le recensement de 1971 nous amène à proposer un certain nombre de corrections au procédé actuel ou tout simplement le rejet de la technique de l'arrondissement aléatoire. Parmi les solutions aux problèmes créés par l'arrondissement aléatoire, mentionnons, dans le cas où on le conserve:

1. L'accès direct des services de diffusion provinciaux aux données du recensement — les services provinciaux contrôlent alors l'arrondissement et les délais de production;
2. La préparation, par Statistique Canada, de bandes d'ordinateur incluant au moins les différents niveaux de découpage géographique les plus utilisés — cette solution suppose

une enquête auprès des utilisateurs —, on éviterait ainsi d'agréger des séries de données déjà arrondies;

3. La mise à la disposition des services de diffusion provinciaux des données brutes du recensement et des programmes de traitement des données y compris celui de l'arrondissement aléatoire.

Il nous semble important que des moyens d'éliminer les ennuis créés par l'arrondissement aléatoire soient étudiés et mis au point avant que les plans de traitement des données du recensement de 1976 soient définitivement arrêtés.

Janvier 1975.