Techniques d'enquête, décembre 1989
Vol. 15, n° 2, pp. 281–290
Statistique Canada

Intervalles de confiance pour les estimations postcensitaires de la population: une étude de cas pour les petites régions

DAVID A. SWANSON[1]

RÉSUMÉ

L'auteur présente une méthode qui permet de construire des intervalles de confiance acceptables pour des estimations postcensitaires de la population en utilisant une version modifiée de la méthode de corrélation des rapports, appelée méthode des rangs. Il montre que le test de Wilcoxon peut servir à déterminer si un modèle de corrélation des rapports donné est stable à long terme. Si la stabilité du modèle est vérifiée, on en conclut que les intervalles de confiance liés aux données qui ont servi à la construction du modèle sont acceptables pour des estimations postcensitaires. Dans le cas contraire, on en conclut que les intervalles de confiance ne sont pas acceptables et qu'en plus, ils sont susceptibles d'exagérer la précision des estimations postcensitaires. Étant donné un modèle instable, l'auteur montre que l'on peut néanmoins déterminer des intervalles de confiance acceptables pour les estimations postcensitaires en utilisant la méthode des rangs. Il présente finalement un exemple empirique où il utilise les estimations de la population des comtés de l'État de Washington.

MOTS CLÉS: Estimation de la population; intervalles de confiance; méthode de régression de la corrélation des rapports.

1. INTRODUCTION

Il n'existait pas de méthode de construction d'intervalles de confiance pour des estimations postcensitaires avant que Espenshade et Tayman (1982) n'élaborent une méthode d'estimation par régression pour séries chronologiques, qui utilise les taux de mortalité postcensitaires par âge. L'introduction de cette méthode représente une étape majeure dans l'évolution des techniques d'estimation; néanmoins, comme la plupart des techniques de pointe, elle présente des lacunes, dont deux méritent d'être soulignées:

1. la méthode risque de ne pas donner de résultats satisfaisants au niveau infra-provincial ou infra-étatique (Espenshade et Tayman 1982);

2. elle s'éloigne beaucoup de la méthode de régression normalement utilisée au Canada et aux États-Unis pour estimer la population de petites régions, c'est-à-dire de la méthode de corrélation des rapports; ce point est très important car il soulève la question de la nature des données qu'il faudrait alors recueillir et de la nouveauté de cette méthode pour les personnes chargées d'estimer la population des petites régions (Statistique Canada 1987). Le terme "petite région" désigne une divison de recensement au Canada (Statistique Canada 1987) et un comté dans presque tous les États américains, les exceptions les plus notables étant l'Alaska, où les petites régions sont définies comme des secteurs de recensement, la Louisiane, où les paroisses sont l'équivalent des comtés, et la Virginie, où les "villes autonomes" sont considérées comme des petites régions.

Dans cet article, nous allons exposer une méthode qui permet de construire des intervalles de confiance pour des estimations postcensitaires de la population de petites régions en utilisant

[1] David A. Swanson, Département de sociologie, Pacific Lutheran University, Tacoma, Washington 98447, U.S.A.

la méthode des rangs, version modifiée de la méthode de corrélation des rapports définie par Swanson (1980), qui applique les principes de la modélisation causale pour tenir compte des changements structurels postcensitaires dans un modèle de corrélation des rapports donné.

La construction d'intervalles de confiance par la méthode de corrélation des rapports soulève trois questions fondamentales. La première concerne la stabilité temporelle du modèle. Si la structure des liens qui existent entre les variables du modèle est fixe dans le temps, les intervalles de confiance construits par rapport à la série de données du modèle s'appliqueront aux estimations de la population produites par le modèle à partir de la série de données d'estimation. Même si de nombreux auteurs s'entendent pour dire qu'il n'est pas prudent de supposer l'invariabilité d'un modèle (D'Allesandro et Tayman 1980; Ericksen 1973, 1974; Mandell et Tayman 1982; Namboodiri 1972; O'Hare 1976, 1980; Smith et Mandell 1984; Spar et Martin 1979; Swanson 1980; Swanson et Prevost 1986; Swanson et Tedrow 1984; Tayman et Schafer 1982; Verma et coll. 1983), il serait utile de disposer d'un test de stabilité. Cela nous amène à la seconde question, soit l'utilisation d'un test statistique. Si le test indique que l'on ne peut supposer la stabilité du modèle et que, malgré cela, on applique, par exemple, des intervalles de confiance fondés sur un modèle construit à l'aide de données de la période 1960-1970 à des estimations produites pour 1979, ces intervalles sont susceptibles d'*exagérer* le niveau de précision des estimations de 1979. D'où la nécessité de trouver une méthode qui produira des intervalles de confiance acceptables. Ce dernier point représente la troisième question relative à la construction d'intervalles de confiance par la méthode de corrélation des rapports.

Dans les sections qui suivent, nous décrivons la méthode de corrélation des rapports ainsi que sa version modifiée à partir de laquelle nous pourrons construire des intervalles de confiance acceptables. Nous décrivons également toute la logique qui sous-tend la construction de ces intervalles de confiance, puis nous passons à un exemple empirique qui illustre le test de stabilité et la production d'intervalles de confiance "satisfaisants" et "insatisfaisants".

2. MÉTHODE D'ESTIMATION DE LA POPULATION

La méthode de corrélation des rapports est une méthode de régression qui sert à mesurer la variation temporelle des proportions de la population de petites régions à partir de la variation observée dans les proportions d'indicateurs symptomatiques comme ceux tirés de la liste électorale, de la liste des cotisants à l'assurance-chômage ou de la liste des effectifs scolaires. On mesure la variation temporelle en calculant simplement un rapport de proportions pour deux périodes différentes.

Comme les chiffres de population pour toutes les petites régions ne peuvent être connus que par le recensement fédéral, on construit toujours un modèle de régression de la corrélation des rapports en utilisant deux périodes séparées par un intervalle régulier. Le modèle est défini par la formule

$$\mathrm{R}_{it} = a_o + \sum_{j=1}^{k} (b_j)\ (X_i)_{jt} + \epsilon$$

où

a_o = ordonnée à l'origine à estimer

b_j = coefficient de régression à estimer

ϵ = terme d'erreur

j = indicateur symptomatique, $(1 \leq j \leq k)$

i = petite région ($1 \leq i \leq n$)

t = année du dernier recensement

et

$$R_{it} = \left[\frac{P_{i,t}}{\sum P_{i,t}}\right] \div \left[\frac{P_{i,t-z}}{\sum P_{i,t-z}}\right] \tag{1.A}$$

$$(X_i)_{t,j} = \left[\frac{S_{i,t}}{\sum S_{i,t}}\right] \div \left[\frac{S_{i,t-z}}{\sum S_{i,t-z}}\right]_j \tag{1.B}$$

où

Z = nombre d'années entre les recensements

P = population

S = indicateur symptomatique.

Une fois le modèle construit, on établit une estimation postcensitaire pour la période $t + x$ en substituant $(S_{i,t+x}/\sum S_{i,t+x})_j$ au numérateur du membre de droite de l'équation [1.B] et $(S_{i,t}/\sum S_{i,t})_j$ au dénominateur. Ainsi, une fois que l'on a déterminé $\hat{R}_{i,t+x}$, on calcule la population réelle de la région i à la période $= t + x$ en introduisant dans l'équation, [1.A] une estimation indépendante du total de la population, P_{t+x}, puis en solutionnant algébriquement cette équation en fonction de $P_{i,t+x}$. Comme $\sum \hat{P}_{i,t+x}$ égale rarement l'estimation indépendante du total de la population P_{t+x}, on effectue une correction de manière à ce que cette estimation corresponde à la somme des chiffres de population.

Une faiblesse de la méthode de corrélation des rapports est l'invariabilité de sa structure. C'est ce qui a amené Swanson (1980) à définir la méthode des rangs. Cette méthode repose sur l'idée, mise de l'avant par Land (1969, Chapitre IV), que l'on peut exploiter l'information contenue dans les coefficients de corrélation totale qui se trouvent dans une série de données d'estimation. L'étude de Land repose sur le théorème fondamental de l'analyse des coefficients de direction énoncé par Wright (1921). Dans cette étude, il est question d'une transformation théorique de la variable dépendante du modèle de régression (variable de population) en une variable explicative non mesurée et d'une structure jusque-là implicite – minimum de trois variables explicatives (dans le modèle de régression), dont la covariance peut être rattachée au fait qu'elles sont elles-mêmes expliquées par la variable de population.

3. MÉTHODE D'ESTIMATION DES INTERVALLES DE CONFIANCE

Si la stabilité des liens observés entre les variables dans la série de données du modèle est vérifiée (comme permet de le faire la méthode des rangs), on devrait retrouver les mêmes liens entre les variables dans la série de données d'estimation. Cette stabilité indique que l'on peut se fonder sur l'erreur type d'estimation (E.T.E.) rattachée à la série de données du modèle pour construire des intervalles de confiance pour la série de données d'estimation. En revanche, si la stabilité n'est pas vérifiée, l'E.T.E. rattachée à la série de données du modèle ne convient pas et peut, de fait, produire des intervalles de confiance qui exagèrent la précision des estimations postcensitaires. Ces considérations nous amènent à nous demander s'il n'y aurait pas lieu de vérifier la stabilité des liens entre les variables par l'inférence statistique.

Pour répondre à cette question, considérons que nous examinons des paires de variables liées entre elles. Cela implique que l'on peut utiliser le test de Wilcoxon pour les observations appariées (Mosteller et Rourke 1973). L'hypothèse nulle dans ce test est qu'il n'y a aucune différence entre les estimations de la population établies à l'aide du modèle de régression normal et celles établies à l'aide du modèle de régression modifié.

Pour bien comprendre la façon dont on construit des intervalles de confiance pour des estimations postcensitaires de la population de petites régions, précisons que la méthode des rangs permet d'obtenir une série de coefficients de régression pour la série de données d'estimation. Dans un deuxième temps, on se sert de ces coefficients pour calculer la valeur estimée de R^2 et de l'E.T.E. pour la série de données d'estimation, puis on se sert finalement de l'E.T.E. estimée pour construire les intervalles de confiance. Rappelons premièrement que le coefficient de détermination multiple, R^2, est simplement la somme des produits du coefficient de corrélation totale entre une variable indépendante et la variable dépendante par le coefficient de régression normalisé pour cette variable indépendante (Hayes 1973), de sorte que l'E.T.E. est définie (Hayes 1973)

$$\text{S.E.E.} = \left[\frac{(n)\ (S_y^2)\ (1 - R^2)}{n - 2}\right]^{1/2}$$

où

n = nombre de cas (petites régions)

S_y^2 = variance de la variable dépendante

R^2 = coefficient de détermination multiple.

Selon Kmenta (1971), la formule de l'intervalle de confiance pour une valeur estimée donnée sur une droite de régression (population) est

$$Y_i \pm (t_{n-2,\alpha/2})\ (\text{S.E.E.}).$$

Il faut bien se rendre compte que l'intervalle de confiance obtenu ne porte pas sur une estimation de la population mais bien sur une estimation d'un rapport de proportions ou R_{it+x}. Or, Espenshade et Tayman (1982) ont montré qu'un intervalle de confiance construit pour une variable pouvait être adapté à une autre variable substituée algébriquement à la première. Par conséquent, en déterminant les limites de confiance de $R_{it+x'}$, nous pouvons les exprimer sous la forme de valeurs de population:

$$(R_{it+x}) \pm (t_{n-2,\alpha/2})\ (\text{S.E.E.})$$

$$= \left[\frac{P_{it+x}}{\sum P_{it+x}}\right] \div \left[\frac{P_{it}}{\sum P_{it}}\right] \pm (t_{n-2,\alpha/2})\ (\text{S.E.E.})$$

ce qui donne

$$\text{L.L.}\ (\hat{P}_{it+x}) =$$

$$\left[\frac{P_{it}}{\sum P_{it}}\right] (\sum P_{it+x}) \left[(\hat{R}_{it+x}) - (t_{n-2,\alpha/2})\ (\text{S.Ê.E.})\right]$$

et

$$\text{U.L. } (\hat{P}_{it+x}) =$$

$$\left[\frac{P_{it}}{\sum P_{it}}\right] (\sum P_{it+x}) \left[(\hat{R}_{it+x}) + (t_{n-2,\alpha/2}) \text{ (S.Ê.E.)}\right]$$

4. ÉTUDE EMPIRIQUE

Le tableau 1.A de Swanson (1980) donne les coefficients de corrélation totale pour un modèle de corrélation des rapports conçu pour estimer la population civile de moins de 65 ans des comtés de l'État de Washington pour la période 1950-1960 et ce, à partir de la liste des cotisants à l'assurance-chômage, de la liste électorale et de la liste des effectifs scolaires pour les huit premiers niveaux. Le tableau 1.B présente les caractéristiques du modèle construit à l'aide de ces données tandis que les tableaux 2.A et 2.B renferment les données correspondantes pour la période 1960-1970. Cette dernière série de données représente les données d'estimation qui serviront à illustrer la méthode.

Bien que nous connaissions toutes les données d'estimation, nous allons appliquer la méthode comme si nous ne les connaissions pas toutes. Évidemment, comme dans tout problème d'estimation, nous connaissons la matrice des coefficients de corrélation totale pour les variables indépendantes, laquelle sert, avec le théorème fondamental de l'analyse des coefficients de direction, à estimer les coefficients du modèle modifié. Si nous appliquons intégralement la méthode des rangs, nous obtenons le modèle modifié (Swanson 1980):

$$Y = 0.046618 + 0.066786X_1 + 0.50727X_2 + 0.38736X_3.$$

Le tableau 1 donne, par comté, l'estimation de la population civile de moins de 65 ans pour 1970 (redressée en fonction de l'estimation indépendante du total de la population de l'État), établie à l'aide du modèle modifié, de même que la population recensée.

Nous avons appliqué le test de Wilcoxon au moyen de la procédure contenue dans la commande "NPAR Tests" du SPSSX (SPSS 1986). Pour des raisons d'économie d'espace, nous ne présenterons ici ni les estimations non modifiées ni les estimations modifiées. On les trouvera dans le tableau 3 de Swanson (1980). Selon l'hypothèse nulle, la probabilité que $Z = -3.2096$ est 0.0013. Nous rejetons donc l'hypothèse nulle et supposons qu'il n'y a pas de continuité pour les comtés de l'État de Washington entre le modèle construit à l'aide des données de la période 1950-1960 et le vrai modèle inconnu rattaché aux données de la période 1960-1970.

Il est intéressant de noter que le test de Chow (Chow 1960) a confirmé les résultats du test de Wilcoxon en montrant que la différence entre le "vrai" modèle de corrélation des rapports (1960-1970) et le modèle pour la période 1950-1960 était statistiquement significative.

Si les résultats du test de Wilcoxon nous avaient amenés à ne pas rejeter l'hypothèse nulle, nous aurions établi les estimations de la population des comtés de l'État de Washington pour 1970 à l'aide des coefficients non modifiés du modèle pour la période 1950-1960. En outre, nous nous serions servis de l'E.T.E. établie pour ce modèle (0.05022) pour construire les intervalles de confiance correspondants. Or, dans ce cas précis, les résultats du test de Wilcoxon nous ont amenés à rejeter l'hypothèse nulle, ce qui indique qu'il faudrait utiliser les coefficients modifiés, déterminés à l'aide de la méthode des rangs, au lieu des coefficients non modifiés. De plus, cela vient confirmer la nécessité de définir une nouvelle E.T.E., qui n'exagérera pas la précision des estimations pour 1970.

Tableau 1

Intervalles de confiance à 90% pour les estimations de la population civile de moins de 65 ans, par comté, État de Washington, 1970

Comté	Population recensée	Limite inférieure	Population estimée	Limite supérieure	Intervalle de confiance à 90% (en pourcentage)
Adams	11102	10335	11458	12581	± 9.80
Asotin	11862	10469	11814	13154	± 11.38
Benton	63144	60405	67511	74616	± 10.53
Chelan	35862	31733	36177	40620	± 12.28
Clallam	30023	28063	31294	34525	± 10.32
Clark	116663	101183	111437	121690	± 9.20
Columbia	3771	3683	4161	4639	± 11.49
Cowlitz	62586	55170	61581	67992	± 10.41
Douglas	15287	14569	16252	17935	± 10.36
Ferry	3336	2963	3397	3831	± 12.78
Franklin	23983	21960	24631	27302	± 10.84
Garfield	2546	2447	2761	3075	± 11.37
Grant	38921	37561	42606	47651	± 11.84
Grays Harbor	52583	46294	52114	57935	± 11.17
Island	20589	20512	22148	24040	± 7.39
Jefferson	9235	8440	9473	10506	± 10.90
King	1054271	935664	1037937	1140203	± 9.85
Kitsap	86529	77022	85821	94619	± 10.25
Kittitas	22764	17649	19863	22077	± 11.15
Klickitat	10729	10440	11923	13406	± 12.44
Lewis	39265	35747	40122	44497	± 10.90
Lincoln	8168	7939	9107	10275	± 12.83
Mason	18411	16057	17827	19596	± 9.93
Okanogan	22952	21002	23795	25688	± 10.97
Pacific	13310	11270	12795	14320	± 11.92
Pend Oreille	5185	5147	5893	6639	± 12.86
Pierce	339048	314272	346728	379184	± 9.36
San Juan	3089	2636	2918	3201	± 9.66
Skagit	45703	43255	48758	54261	± 11.29
Skamania	5330	4787	5358	5929	± 10.66
Snohomish	245193	213164	231996	250827	± 8.12
Spokane	251057	227372	256723	286072	± 11.43
Stevens	15178	13869	15780	17692	± 12.11
Thurston	68719	63644	69540	75436	± 8.48
Wahkiakum	3137	3033	3397	3761	± 10.72
Walla Walla	36608	33727	38271	42812	± 11.87
Whatcom	72111	63218	70670	78122	± 10.54
Whitman	34843	28960	32409	35858	± 10.64
Yakima	128960	120347	136203	152219	± 11.69

En utilisant les estimations calculées dans l'exemple de l'État de Washington (Swanson 1980), nous obtenons

$$\hat{R}^2 = (0.07533)\,(0.75290) + (0.47085)\,(0.92146) + (0.49481)\,(0.88082) = 0.926$$

et

$$(\text{S.Ê.E.}) = \left[\frac{(39)\,(0.2145)^2\,(1 - 0.926)}{39\text{-}2}\right]^{1/2}$$

$$= 0.0599.$$

Notons que, d'après le tableau 2 de Swanson (1980), la valeur réelle du coefficient de détermination et de l'E.T.E. est 0.878 et 0.05077 respectivement. L'E.T.E. estimée est donc supérieure à l'E.T.E. réelle. Cela est normal étant donné que nous sommes moins sûrs de la précision des estimations établies par la méthode des rangs que de celle des estimations établies à l'aide du "vrai" modèle, si, bien sûr, ce modèle peut être déterminé. Par la méthode des rangs, nous pouvons maintenant construire une bande de confiance à l'aide de la formule suivante:

$$Y_i \pm (t_{37,\alpha/2})\,(0.0599).$$

Le tableau 1 donne également les limites d'intervalle de confiance à 90 % pour les estimations de la population de comté de 1970. Ces intervalles de confiance sont définis par la formule:

$$\left[\frac{P_{i1960}}{2522141}\right](3032053)\left[(\hat{R}_{i1970}) \pm (1.69)\,(0.0599)\right].$$

Lorsqu'on compare les intervalles de confiance du tableau 1 à la population recensée correspondante, on s'aperçoit que dans un seul cas (comté de Kittitas), la population recensée n'est pas incluse dans l'intervalle à 90 %. De fait, elle excède de 687 la limite supérieure de l'intervalle. À un niveau de confiance de 90 %, les intervalles sont relativement larges: moyenne de ±10.81 %, minimum de ±7.39 % (comté d'Island) et maximum de ±12.83 % (comté de Lincoln). Si on compare ces chiffres à l'écart moyen en pourcentage pour les estimations de 1970, qui est de 4.89 (Swanson 1980), on en déduit qu'un seuil de 90 % donne des intervalles trop étendus au point de vue pratique. On aurait donc avantage à choisir un niveau de confiance plus approprié. On aurait aussi avantage à étudier les conséquences de l'utilisation de l'E.T.E. non modifiée (0.05022) tirée du modèle pour la période 1950-1960. On s'attend que le modèle non modifié produise des intervalles de confiance trop "optimistes", c'est-à-dire qu'à un niveau de confiance donné, le nombre de comtés dont la population réelle se trouve dans l'intervalle soit moins élevé que prévu. Pour approfondir ces questions, nous avons construit le tableau 2.

Le tableau 2 renferme deux séries de données distinctes. Dans les deux cas toutefois, il s'agit d'une comparaison entre les estimations d'un modèle modifié et d'un modèle non modifié et entre les intervalles de confiance correspondants. En ce qui a trait à la question des intervalles de confiance optimistes pour les estimations de 1970 établies à l'aide du modèle non modifié, le tableau 2 indique que pour des niveaux de confiance allant de 50 à 90 %, les intervalles sont effectivement optimistes puisqu'il n'y a que deux niveaux de confiance pour lesquels le nombre de comtés est conforme au niveau de précision voulu. À un seuil de 80 %, par exemple, l'intervalle contient la population recensée de 28 comtés seulement (72 %); à un seuil de 60 %, l'intervalle contient la population recensée de 22 comtés seulement (56 %).

Tableau 2

Nombre (%) de comtés dont la population recensée en 1970 est incluse dans l'intervalle de confiance

Niveau de confiance	E.T.E. non modifiée (0.05022)	E.T.E. modifiée (0.0599)
90%	35 (89.7%)	38 (97.4%)
80%	28 (71.8%)	33 (84.6%)
70%	24 (61.5%)	29 (80.6%)
66.66%	24 (61.5%)	26 (66.66%)
60%	22 (56.4%)	23 (59.0%)
50%	20 (51.3%)	22 (56.4%)
	Intervalle moyen (en pourcentage)	
	E.T.E. non modifiée (0.05022)	E.T.E. modifiée (0.0599)
90%	9.10	10.81
80%	7.02	8.38
70%	5.66	6.75
66.66%	5.59	6.40
60%	4.59	5.47
50%	3.66	4.35

Le second aspect du tableau 2 est l'intervalle moyen qui correspond à un niveau de confiance donné. À un seuil de 90 %, l'intervalle moyen est de 9.10 % selon le modèle non modifié et de 10.81 % selon le modèle modifié. À un seuil de 50 %, il est de 3.66 % et de 4.35 % respectivement. De toute évidence, avec des niveaux de confiance de 50 et de 60 %, on obtient des intervalles moyens qui sont plus en rapport avec l'écart moyen en pourcentage, qui est de 4.88 pour le modèle modifié.

Lorsqu'on étudie la question des intervalles de confiance, on sent la nécessité d'élaborer une méthode de construction d'intervalles de confiance qui produira des intervalles fiables en ce qui a trait à la précision des estimations postcensitaires de la population de petites régions. En revanche, il faut aussi savoir choisir un niveau de confiance conforme aux estimations. Soulignons à cet égard l'étude de Stoto (1983) sur les intervalles de confiance empiriques pour des projections démographiques. Stoto (1983) a constaté notamment que les projections démographiques les plus élevées et les plus faibles produites par le U.S. Bureau of the Census (1977) pour les États-Unis correspondaient à un intervalle de confiance à 66.66 %. Il se peut que ce niveau convienne pour des estimations postcensitaires de la population de petites régions même si, dans le test que nous avons effectué, il produit un intervalle moyen de 6.4 % selon le modèle modifié, ce qui est quelque peu supérieur à l'écart moyen (4.9 %). Il y a aussi la question du délai entre l'année pour laquelle on veut établir des estimations postcensitaires et l'année du dernier recensement. Dans l'exemple que nous venons d'étudier, on a utilisé la période postcensitaire maximum prévue aux États-Unis, soit 10 ans. Essentiellement, nous sommes devant

une situation où la fiabilité des estimations est à son plus faible niveau pour chaque comté. Dans ces circonstances, l'intervalle relativement large que l'on obtient pour chaque comté à un seuil de 90 % n'est pas exagéré. Le modèle pourrait subir des changements structurels au fil du temps. Ainsi, nous aurions probablement une bande de confiance qui irait en s'élargissant tout le long de la période intercensitaire qui suit l'année du dernier recensement utilisée dans la construction du modèle.

5. CONCLUSION

Il est entendu que cet article ne vise pas à présenter la méthode des rangs comme un moyen infaillible de construire des intervalles de confiance pour des estimations postcensitaires de la population de petites régions. Cette méthode semble toutefois représenter une bonne base de départ pour la recherche. Malgré les faiblesses relevées, les personnes chargées d'établir des estimations postcensitaires de la population devraient pouvoir tirer profit de l'utilisation du test de Wilcoxon et des intervalles de confiance construits à l'aide de la méthode des rangs. Tout d'abord, il est important, comme le soulignent Espenshade et Tayman (1983), que les utilisateurs d'estimations postcensitaires de la population aient une idée de la précision de ces estimations et c'est justement là l'objet du test de Wilcoxon et des intervalles de confiance. Deuxièmement, grâce à la construction d'intervalles de confiance acceptables, on dispose d'un moyen formel pour résoudre tout désaccord concernant la population d'une petite région donnée en ayant recours à des tests d'hypothèses. Troisièmement, l'E.T.E. peut servir de critère pour le choix d'un modèle. Cela signifie qu'il est possible d'envisager une série de modèles de corrélation des rapports pour n'importe quelle année d'estimation postcensitaire donnée et qu'en outre, on dispose d'un critère formel pour le choix du modèle. Ces éléments pourraient être utiles lorsque les estimations produites à l'aide de la méthode de corrélation des rapports par le service de la démographie d'une administration publique sont contestées pour une année postcensitaire donnée, phénomène devenu plus fréquent, particulièrement aux É.-U. (D'Allesandro 1987).

REMERCIEMENTS

L'auteur tient à remercier Jeff Tayman ainsi que des arbitres anonymes et le comité de rédaction pour leurs commentaires et suggestions. Il tient aussi à exprimer sa reconnaissance à Peggy Jobe, qui a tapé la version préliminaire de cet article, et à Carey Taylor, qui a tapé la version finale.

BIBLIOGRAPHIE

CHOW, G. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrics*, 28, 591-605.

D'ALLESANDRO, F. (1987). Should applied demographers take out liability insurance? Article présenté au Annual Meeting of The Population Association of America.

D'ALLESANDRO, F., et TAYMAN, J. (1980). Ridge regression for population estimation: Some insights and clarification. Staff Document No. 56. Office of Financial Management, State of Washington: Olympia, Washington.

DRAPER, N.R., et SMITH, H. (1981). *Applied Regression Analysis, 2nd Edition*. New York: Wiley.

ERICKSEN, E.P. (1973). A method for combining sample survey data and symptomatic indicators to obtain population estimates for local areas. *Demography*, 10, 137-160.

ERICKSEN, E.P. (1974). A regression method for estimating population changes of local areas. *Journal of the American Statistical Association*, 69, 867- 875.

ESPENSHADE, T.J., et TAYMAN, J. (1982). Confidence intervals for postcensal state population estimates. *Demography*, 19, 191-210.

HAYS, W.L. (1973). *Statistics for the Social Sciences*. New York: Holt, Rinehart and Winston.

KMENTA, J. (1971). *Elements of Econometrics*. New York: Macmillan.

LAND, K.C. (1969). Explorations in mathematical sociology. Thèse de doctorat. University of Texas, Austin.

MANDELL, M., et TAYMAN, J. (1982). Measuring temporal stability in regression models of population estimation. *Demography*, 19, 1351-46.

MOSTELLER, F., et ROURKE, R. (1973). *Sturdy Statistics*. Reading, Massachusetts: Addison-Wesley.

NAMBOODIRI, N.K. (1972). On the ratio-correlation and related methods of subnational population estimation. *Demography*, 9, 443-453.

O'HARE, W. (1976). Report on a multiple regression method for making population estimates. *Demography*, 13, 369-379.

O'HARE, W. (1980). A note on the use of regression methods in population estimates. *Demography*, 17, 341-343.

SMITH, S., et MANDELL, M. (1984). A comparison of local population estimates: The housing unit method versus component II, regression, and administrative records. *Journal of the American Statistical Association*, 99, 292-289.

SPAR, M., et MARTIN, J. (1979). Refinements to regression-based estimates of postcensal population characteristics. *Review of Public Data Use*, 7, 16-22.

SPSS, Inc. (1986). *SPSSx User's Guide*. Chicago: SPSS, Inc.

STATISTIQUE CANADA (1987). *Méthodes d'estimation de la population, Canada*. N° 91-528F au catalogue, Statistique Canada.

STOTO, M.A. (1983). The accuracy of population projections. *Journal of the American Statistical Association*, 78, 13-20.

SWANSON, D. (1980). Improving accuracy in multiple regression estimates of population using principles from causal modeling. *Demography*, 17, 413-427.

SWANSON, D., et PREVOST, R. (1986). Identifying extreme errors in ratio-correlation estimates of population. Article présenté au Annual Meeting of the Population Association of America.

SWANSON, D., et TEDROW, L. (1984). Improving the measurement of temporal change in regression models used for county population estimates. *Demography*, 21, 373-381.

TAYMAN, J., et SCHAFER, E. (1985). The impact of coefficient drift and measurement error on the accuracy of ratio-correlation population estimates. *The Review of Regional Studies*, 15, 3-10.

U.S. BUREAU OF THE CENSUS (1977). Projections of the Population of the United States, 1977 to 2050. *Current Population Reports. Series P-25 No. 704*. Washington, D.C.: U.S. Government Printing Office.

VERMA, R.V.P., BASAVARAJAPPA, K.G., et BENDER, R.K. (1983). Estimation par regression de la population à l'échelon infraprovincial au Canada. *Techniques d'enquête*, 9, 242-266.

WRIGHT, S. (1921). Correlation and causation. *Journal of Agricultural Research*, 20, 557-585.