Techniques d'enquête, juin 1990
Vol. 16, n° 1, pp. 155–161
Statistique Canada

Un exemple d'utilisation de tests aléatoires pour les essais du questionnaire du recensement

YVES BÉLAND et ALAIN THÉBERGE[1]

RÉSUMÉ

L'Essai modulaire 2, une enquête de Statistique Canada dont le but était d'aider à la mise au point du questionnaire du recensement de 1991 utilisait deux questionnaires distincts. L'échantillon de l'enquête était non probabiliste. Cet article décrit brièvement la méthodologie de l'enquête et comment les tests aléatoires ont été utilisés pour comparer les deux questionnaires.

MOTS CLÉS: Tests aléatoires; échantillonnage non probabiliste; plan d'expérience.

1. INTRODUCTION

On peut distinguer deux types de tests statistiques. Il y a ceux pour lesquels la décision est basée sur une comparaison de la valeur observée d'une statistique avec la distribution, sous l'hypothèse nulle, des valeurs de la même statistique pour l'ensemble des échantillons qui auraient pu être tirés. On les appellera les tests classiques. Pour faire un tel test, il faut connaître la probabilité d'obtenir un échantillon quelconque. Il doit donc y avoir eu échantillonnage probabiliste selon un plan de sondage connu. D'autre part, il y a les tests pour lesquels la décision est basée sur une comparaison de la valeur observée d'une statistique avec la distribution, sous l'hypothèse nulle, des valeurs de la même statistique pour l'ensemble des permutations possibles des données. On les appellera les tests aléatoires. Fisher (1935) en a fait l'utilisation pour comparer deux échantillons de graines, et Edgington (1987) traite de plusieurs aspects de ces tests. Les tests aléatoires requièrent d'abord l'existence de "traitements" afin de définir les permutations, ensuite il faut connaître la probabilité d'obtenir une permutation quelconque. On doit donc avoir choisi de façon aléatoire quelle unité allait être soumise à quel traitement, c'est-à-dire qu'il faut un plan d'expérience avec randomisation. Dans un organisme comme Statistique Canada, ce sont habituellement des tests classiques qui sont employés. Cet état de choses reflète d'une part, le fait que la plupart des enquêtes-échantillon effectuées à Statistique Canada utilisent un échantillonnage probabiliste, et d'autre part, l'absence de "traitements" dans ces enquêtes. On décrira dans cet article une enquête faisant exception à la règle, et comment on a appliqué les tests aléatoires.

Dans la section deux, on décrit brièvement la méthodologie employée pour les essais modulaires. La troisième section décrit à l'aide d'exemples simples la procédure utilisée pour un test aléatoire. L'application des tests aléatoires à l'Essai modulaire 2 est présentée dans la section quatre.

2. ESSAIS MODULAIRES

Dans le cadre de la planification du recensement de 1991, deux essais modulaires ont été effectués dans le but de valider les questions qui seront susceptibles d'être posées lors de ce recensement. Ces enquêtes visent à s'assurer que le libellé de chaque question nouvelle ou

[1] Yves Béland, Division des méthodes d'enquêtes sociales, Alain Théberge, Division des méthodes d'enquêtes-entreprises, Statistique Canada, Parc Tunney, Ottawa (Ontario), Canada, K1A 0T6.

reformulée, ne pose pas de problèmes de compréhension auprès des répondants. On parle d'essais modulaires car ces enquêtes sont indépendantes les unes des autres et portent sur des parties distinctes du contenu du questionnaire du recensement.

Le premier essai, qui s'est tenu au mois de novembre 1987, avait comme objectif de mettre au point de nouvelles questions portant sur la couverture de la population, l'état matrimonial, la fécondité, le travail bénévole ainsi que sur la nuptialité. Cette première enquête n'a fait l'objet d'aucun test, classique ou aléatoire.

Le deuxième essai, qui s'est tenu au mois de janvier 1988, visait principalement à mesurer la réaction des populations ethniques face à certaines questions portant sur la langue, l'origine ethnique, la religion, la citoyenneté ainsi que sur la mobilité. L'échantillon d'environ 3,500 ménages du deuxième essai modulaire provient d'un échantillonnage a deux degrés effectué à l'intérieur des régions métropolitaines de Halifax, Québec, Montréal, Toronto, Winnipeg et Vancouver. Pour des raisons de coût et pour faciliter la collecte des données, mais aussi parce qu'on voulait que l'échantillon contienne plusieurs personnes d'origines ethniques diverses, une méthode de sélection non probabiliste a été retenue. Le questionnaire du deuxième essai modulaire existe en deux versions dont les différences sont décrites à la section 4. Les ménages sélectionnés dans l'échantillon ont reçu aléatoirement un ou l'autre de ces questionnaires.

Pour nous permettre de tester statistiquement certaines hypothèses quant à ce deuxième essai modulaire, on a utilisé la méthode des tests aléatoires. Les tests aléatoires peuvent être utilisés pour comparer deux traitements appliqués aux unités d'échantillons qui ne sont pas nécessairement probabilistes.

3. TESTS ALÉATOIRES

Voici la procédure utilisée pour un test aléatoire. On calcule la valeur d'une statistique pour les données observées, puis on calcule la valeur de la même statistique pour les autres permutations des données qui sont possibles avec le plan d'expérience utilisé. On rejette H_0, si on juge que la valeur de la statistique pour les données observées est extrême par rapport aux valeurs obtenues sous H_0 pour l'ensemble des permutations.

Par exemple, supposons qu'il y a quatre ménages: le premier ménage est formé de trois personnes, les ménages deux et trois comptent deux personnes chacun, et le quatrième ménage en compte une. Ces ménages peuvent avoir été sélectionnés arbitrairement, mais on choisit de façon aléatoire un ménage aux membres duquel on applique le traitement Y, et on applique le traitement X aux membres des trois autres. Disons que le quatrième ménage est choisi pour le traitement Y, que pour le premier ménage le traitement a réussi pour deux des trois personnes, que pour les ménages deux et trois le traitement a réussi pour une des deux personnes, tandis que pour la personne du quatrième ménage le traitement n'a pas réussi. Notre hypothèse nulle est que les résultats sont indépendants du traitement employé. Pour mesurer l'impact du traitement X par rapport au traitement Y, on calcule la statistique S qui représente le nombre moyen de réussites pour le traitement X moins le nombre moyen de réussites pour le traitement Y. Dans notre cas $S = (2 + 1 + 1)/(3 + 2 + 2) - 0/1 = 4/7$. Pour savoir si cette valeur est significative, on présente dans le tableau 1 les valeurs de S obtenues en permutant les observations. Ici, une observation est constituée du nombre de personnes dans le ménage (après la barre verticale dans le tableau) et du nombre de réussites dans le ménage (devant la barre verticale dans le tableau). Si on fait un test à droite, on rejette H_0 si $\alpha \geq 3/12 = .25$, puisque trois des douze permutations donnent une valeur de S supérieure ou égale à 4/7, la valeur observée.

Plutôt que de permuter les observations, on aurait pu permuter les traitements. Le tableau 2 montre ce qu'on obtient dans ce cas. Encore une fois, pour un test à droite, on rejette H_0 si $\alpha \geq 1/4 = .25$, puisqu'une seule des quatre permutations donne une valeur de S supérieure ou égale à 4/7. Ce n'est pas une coïncidence si le résultat est le même. En effet, notons n_{ki} le

Tableau 1

Valeurs de la statistique S pour chaque permutation des observations

Traitement	Permutations											
X	2 | 3	1 | 2	1 | 2	2 | 3	2 | 3	1 | 2	1 | 2	0 | 1	0 | 1	1 | 2	1 | 2	0 | 1
X	1 | 2	2 | 3	1 | 2	1 | 2	0 | 1	2 | 3	0 | 1	1 | 2	2 | 3	1 | 2	0 | 1	1 | 2
X	1 | 2	1 | 2	2 | 3	0 | 1	1 | 2	0 | 1	2 | 3	2 | 3	1 | 2	0 | 1	1 | 2	1 | 2
Y	0 | 1	0 | 1	0 | 1	1 | 2	1 | 2	1 | 2	1 | 2	1 | 2	1 | 2	2 | 3	2 | 3	2 | 3
S	4/7	4/7	4/7	0	0	0	0	0	0	−4/15	−4/15	−4/15

Tableau 2

Valeurs de la statistique S pour chaque permutation des traitements

Observation	Permutations			
2 | 3	X	X	X	Y
1 | 2	X	X	Y	X
1 | 2	X	Y	X	X
0 | 1	Y	X	X	X
S	4/7	0	0	−4/15

nombre d'unités qui reçoivent le traitement k ($k = 1, \ldots, K$) et pour lesquelles on observe le résultat r_i ($i = 1, \ldots, I$), $n_{k.} = \sum_i n_{ki}$ le nombre d'unités recevant le traitement k, $n_{.i} = \sum_k n_{ki}$ le nombre d'unités pour lesquelles on observe le résultat r_i et $n_{..} = \sum_k \sum_i n_{ki}$ le nombre total d'unités. Le nombre, N_t, de permutations des traitements est donné par

$$N_t = n_{..}! \Big/ \prod_k (n_{k.}!). \tag{1}$$

De ces N_t permutations, il y en a N_t^* pour lesquelles il y a n_{ki} unités associées au traitement k et au résultat r_i ($k = 1, 2, \ldots, K$; $i = 1, 2, \ldots, I$), où

$$N_t^* = \prod_i \left(n_{.i}! \Big/ \prod_k (n_{ki}!) \right). \tag{2}$$

D'autre part, il y a N_o permutations des observations, où

$$N_o = n_{..}! \Big/ \prod_i (n_{.i}!). \tag{3}$$

De ces N_o permutations, il y en a N_o^* pour lesquelles on a n_{ki} unités associées au traitement k et au résultat r_i ($k = 1, 2, \ldots, K$; $i = 1, 2, \ldots, I$), où

$$N_o^* = \prod_k \left(n_{k.}! \Big/ \prod_i (n_{ki}!) \right). \tag{4}$$

Puisque $N_o^*/N_o = N_t^*/N_t$, les tests sont équivalents. Afin de minimiser les calculs, il vaut mieux permuter les traitements si $N_t < N_o$, et permuter les observations si $N_t > N_o$. Lorsque le nombre de permutations est grand, il a été suggéré par Dwass (1957) de tirer un échantillon de ces permutations et comparer la valeur observée de la statistique à l'ensemble des valeurs pour l'échantillon. Le fait de ne pas considérer toutes les permutations n'affecte pas le niveau du test mais seulement sa puissance. Si on échantillonne les permutations, la règle donnée plus haut tient toujours, non plus afin de minimiser les calculs, mais afin de minimiser la perte de puissance due à l'échantillonnage. Dwass montre, par exemple, que pour un test unilatéral au niveau 0.05, la perte de puissance avec un échantillon de 999 permutations est d'au plus 5.5%. Tel que mentionné dans Bradley (1968), les résultats d'une comparaison entre les tests aléatoires et les tests classiques quant à la puissance dépendront de la mesure dans laquelle les présupposés des tests classiques sont satisfaits.

À cause de la façon dont sont construits les tests aléatoires, l'inférence ne porte que sur l'efficacité du traitement sur les unités de l'échantillon et non sur la population entière. D'un autre côté, les tests classiques sont basés sur un échantillon probabiliste tiré d'une population qui est rarement celle qui nous intéresse. Dans le cas qui nous occupe, par exemple, la population d'intérêt est la population canadienne le jour du recensement, le 4 juin 1991. Ainsi, pour les deux types de tests, des arguments non statistiques doivent être invoqués pour généraliser les inférences à la population d'intérêt.

4. UTILISATION DES TESTS ALÉATOIRES POUR L'ESSAI MODULAIRE 2

Comme il a été mentionné précédemment, le questionnaire du deuxième essai modulaire existe en deux versions: les versions «X» et «Y». La différence se situe au niveau des questions sur l'origine et l'identité ethnique. En effet, dans la version «X», «CANADIEN» constitue une réponse possible (une case est réservée à cette fin) pour ces questions. La version «Y» n'offre pas ce choix de réponse pour ces mêmes questions, les personnes désirant répondre «CANADIEN» doivent inscrire cette réponse en toutes lettres après avoir coché la case «AUTRE(S)».

On désirait savoir si ces questions du questionnaire d'essai amènent plus ou moins de réponses multiples dans la version «X» que dans la version «Y». On dit qu'il y a une réponse multiple lorsque plus d'un choix de réponse est indiqué. On voulait aussi connaître l'influence du type de questionnaire sur la multiplicité (nombre de choix de réponse indiqués par le répondant) et sur certains choix de réponse (comme «FRANÇAIS») pour ces questions. Les types de questionnaire constituent les traitements. Parce que l'échantillon de chaque région avait ses particularités, les tests aléatoires ont été effectués séparément pour chacune des régions métropolitaines parmi lesquelles l'échantillon fut sélectionné.

Pour débuter, on génère aléatoirement un échantillon de 999 permutations de types de questionnaire. Une permutation est générée de la manière suivante. Pour une région donnée, soient N_x et N_y le nombre de questionnaires «X» et «Y» respectivement. En utilisant l'algorithme de Bebbington (1975), on choisit parmi les $N_x + N_y$ ménages un échantillon aléatoire simple de N_x ménages auxquels on assigne le questionnaire de type «X». Ce processus est répété 999 fois. Par la suite, pour une question donnée, on calcule la proportion d'individus ayant donné une réponse multiple pour chacune des versions «X» et «Y» qu'on dénote P_x et P_y.

Ensuite, pour chacune des 999 permutations de types de questionnaire ainsi que pour l'échantillon initial observé, on calcule la statistique $S = P_x - P_y$. On se retrouve donc avec 1000 valeurs de S que l'on classe en ordre croissant. Si, lors de cette classification, deux ou plusieurs statistiques sont égales, un nombre aléatoire entre 0 et 1 est généré et c'est ce nombre qui déterminera l'ordre parmi les statistiques égales. La variable RANGP_{x-y} représente le rang de la statistique S observée.

Soient μ_x et μ_y les espérances de la proportion d'individus donnant une réponse multiple pour les versions «X» et «Y» respectivement. Pour toutes les régions, excluant celle de Halifax, on teste:

$$H_0: \mu_x = \mu_y$$

versus

$$H_1: \mu_x > \mu_y.$$

Pour Halifax, on utilise la contre-hypothèse $H_1: \mu_x < \mu_y$. La raison pour laquelle on utilise une contre-hypothèse différente pour Halifax comparativement aux autres régions est que la majorité des ménages qui ont été sélectionnés dans cette région sont composés de personnes de souche britannique et que pour cette raison, on s'attend à avoir plus de réponses multiples avec le questionnaire «Y». Dans le questionnaire de type «Y», «CANADIEN» n'est pas un choix de réponse pour les questions sur l'identité et l'origine ethnique, et comme la plupart des personnes de cette région sont d'origine anglaise, écossaise ou irlandaise, ces gens qui ont reçu un questionnaire «Y» ont coché une ou plusieurs de ces cases. Pour ce qui est des ménages qui ont reçu un questionnaire «X», les personnes de ces ménages peuvent tout simplement coché la case «CANADIEN».

Le niveau critique, $\hat{\alpha}$, est calculé de la manière suivante: pour la région de Halifax, étant donné qu'on rejette H_0 si la proportion chez les «X» est significativement inférieure a celle observée chez les «Y», le niveau critique est $\text{RANGP}_{x-y}/1000$. Par contre, pour toutes les autres régions, étant donné qu'on rejette H_0 si la proportion chez les «X» est significativement supérieure a celle observée chez les «Y», le niveau critique est $(1001\text{-RANGP}_{x-y})/1000$. Les résultats sont présentés dans le tableau 3.

La méthode des tests aléatoires a également été utilisée pour tester la multiplicité, c'est-à-dire le nombre de cases cochées par le répondant, pour les questions sur l'origine et l'identité ethnique dans chacune des régions. Cette fois-ci, au lieu de parler de proportions (P_x, P_y), on parle de ratios (R_x, R_y). Le ratio R_x (R_y) est le nombre moyen de cases cochées par répondant pour une question des questionnaires «X» «Y». Par la suite, la méthode demeure la même et au lieu de parler de RANGP_{x-y}, on parle de RANGR_{x-y} et la statistique S se définit maintenant comme étant $R_x - R_y$. Cependant, étant donné la plus grande variabilité des statistiques S pour les tests quant à la multiplicité, on a généré un échantillon aléatoire de 1999 permutations au lieu de 999.

Soient F et G les fonctions de répartition du nombre de cases cochées pour un questionnaire «X» et un questionnaire «Y» respectivement. Pour toutes les régions, excluant celle de Halifax, on teste l'hypothèse:

$$H_0: F = G$$

versus

$$H_1: F(z) \leq G(z) \text{ pour tout } z \text{ et } F \neq G.$$

Si on rejette H_0, on dit alors que le nombre de cases cochées pour un questionnaire «X» est stochastiquement plus grand que le nombre de cases cochées pour un questionnaire «Y». Pour Halifax, on utilise la contre-hypothèse $H_1: F_{(z)} \geq G_{(z)}$ pour tout z et $F \neq G$. Les résultats sont présentés dans le tableau 3. On peut remarquer que que la valeur de R_y dans la région de Québec pour chacune des questions est inférieure à 1. La raison est qu'à Québec, la plupart des gens n'ont coché qu'une case et que certains répondants n'ont pas répondu à l'une ou l'autre des questions.

Pour terminer, on a comparé les versions «X» et «Y» du deuxième essai modulaire quant au nombre de personnes s'identifiant comme d'origine ou d'identité française, italienne ou britannique dans certaines régions. Par britannique, on entend qu'au moins une des cases

Tableau 3
Niveaux critiques pour le taux de réponse multiple et la multiplité

		Multiple			Multiplicité		
Question	Région	P_x	P_y	$\hat{\alpha}$	R_x	R_y	$\hat{\alpha}$
ORIGINE	HALIFAX	0.435	0.536	0.087	1.617	1.914	0.062
ORIGINE	QUÉBEC	0.154	0.043	0.001	1.143	0.986	0.001
ORIGINE	MONTRÉAL	0.185	0.194	0.612	1.141	1.152	0.585
ORIGINE	TORONTO	0.127	0.122	0.393	1.124	1.125	0.495
ORIGINE	WINNIPEG	0.293	0.307	0.622	1.439	1.398	0.345
ORIGINE	VANCOUVER	0.285	0.296	0.621	1.440	1.392	0.280
IDENTITÉ	HALIFAX	0.220	0.335	0.035	1.244	1.502	0.029
IDENTITÉ	QUÉBEC	0.140	0.016	0.001	1.131	0.959	0.001
IDENTITÉ	MONTRÉAL	0.159	0.125	0.063	1.075	1.044	0.186
IDENTITÉ	TORONTO	0.186	0.120	0.001	1.154	1.075	0.005
IDENTITÉ	WINNIPEG	0.224	0.195	0.248	1.253	1.208	0.298
IDENTITÉ	VANCOUVER	0.186	0.183	0.457	1.182	1.137	0.202

Tableau 4
Niveaux critiques pour certaines variables

Question	Variable	Région	P_x	P_y	$\hat{\alpha}$
ORIGINE	FRANÇAIS	QUÉBEC	0.127	0.897	0.001
ORIGINE	FRANÇAIS	MONTRÉAL	0.038	0.210	0.001
ORIGINE	BRITANNIQUE	HALIFAX	0.321	0.837	0.001
ORIGINE	BRITANNIQUE	MONTRÉAL	0.034	0.092	0.002
ORIGINE	BRITANNIQUE	TORONTO	0.085	0.135	0.003
ORIGINE	BRITANNIQUE	WINNIPEG	0.167	0.234	0.054
ORIGINE	BRITANNIQUE	VANCOUVER	0.267	0.325	0.065
IDENTITÉ	FRANÇAIS	QUÉBEC	0.138	0.899	0.001
IDENTITÉ	BRITANNIQUE	HALIFAX	0.153	0.828	0.001
IDENTITÉ	BRITANNIQUE	MONTRÉAL	0.022	0.117	0.001
IDENTITÉ	BRITANNIQUE	TORONTO	0.050	0.215	0.001
IDENTITÉ	BRITANNIQUE	WINNIPEG	0.074	0.276	0.001
IDENTITÉ	BRITANNIQUE	VANCOUVER	0.104	0.325	0.001
IDENTITÉ	ITALIEN	TORONTO	0.412	0.463	0.060

«IRLANDAIS», «ÉCOSSAIS» ou «ANGLAIS» a été cochée. Par exemple, si on veut faire un test quant à la proportion de gens qui ont répondu «FRANÇAIS», on définit μ_x et μ_y comme étant les espérances de la proportion de questionnaires avec la réponse «FRANÇAIS» pour les versions «X» et «Y». Pour toutes les régions, on teste:

$$H_0: \quad \mu_x = \mu_y$$

versus

$$H_1: \quad \mu_x < \mu_y.$$

La méthode des tests aléatoires a été utilisée avec 999 permutations. Les résultats sont donnés dans le tableau 4.

5. CONCLUSION

On remarque que les résultats des tests pour le taux de réponse multiple sont similaires à ceux pour la multiplicité, ce qui n'est pas surprenant. En comparant les niveaux critiques pour la question sur l'origine ethnique à ceux pour l'identité ethnique, on voit que ce sont les réponses à la question sur l'identité ethnique qui sont le plus affectées par la différence entre les deux versions de questionnaires.

La principale raison pour laquelle on a eu recours aux tests aléatoires est que l'échantillon de l'Essai modulaire 2 n'était pas probabiliste. Il y a cependant d'autres cas où l'emploi de ces tests est approprié. Par exemple, pour un test t de Student de l'égalité des moyennes, on doit avoir recours à l'hypothèse de normalité et on suppose également qu'il y a égalité des variances. Ces présupposés sont inutiles pour un test aléatoire. Il faut toutefois se rappeler que les conclusions qu'on peut tirer à partir des tests aléatoires s'appliquent à l'échantillon et pas nécessairement à la population entière, à moins qu'il ne s'agisse d'un échantillon aléatoire simple.

BIBLIOGRAPHIE

BEBBINGTON, A.C. (1975). A simple method of drawing a sample without replacement. *Applied Statistics*, 24, 136.

BRADLEY, J.V. (1968). *Distribution-free Statistical Tests*. Englewood Cliffs: Prentice-Hall.

DWASS, M. (1957). Modified randomization tests for nonparametric hypotheses. *Annals of Mathematical Statistics*, 28, 181-187.

EDGINGTON, E.S. (1987). *Randomization Tests*, (2e éd.). New York: Marcel Dekker.

FISHER, R.A. (1935). *The Design of Experiments*. Edinburgh: Oliver and Boyd.