Techniques d'enquête, juin 1978
Vol. 4, nº 1
Statistique Canada, nº 12-001-X au catalogue

L'effet d'un plan de sondage à deux degrés sur les tests d'indépendance dans un tableau 2 par 2

J. Cowan et D.A. Binder[1]

Résumé

Quand on utilise un échantillon à deux degrés pour rassembler des données, les corrélations entre les unités échantillonnées rendent le test d'indépendance $\chi 2$ invalide. Si on utilise les tables ordinaires de $\chi 2$, on obtient généralement un test qui est plus grand que le seuil significatif voulu. L'effet du plan d'échantillonnage provient de deux facteurs principaux : la taille de l'échantillon dans les UPE et le degré d'indépendance des caractéristiques dans chaque UPE. L'effet de la taille de l'échantillon dans les UPE est à son maximum quand il n'y a pas d'indépendance dans chaque UPE, et diminue à mesure que le degré d'indépendance augmente.

Mots-clés : Échantillonnage à deux degrés; test d'indépendance; distribution de Dirichlet.

1 Introduction

L'inférence statistique classique s'est développée au fil des ans en posant l'hypothèse d'une indépendance des observations. Ces dernières années, l'attention s'est tournée vers les tentatives d'élaboration de techniques d'analyse des données pour des procédures d'échantillonnage complexe, en particulier dans le domaine des enquêtes sociales (Kish et Frankel [5]). Dans le cadre de l'Enquête santé Canada, un groupe d'analyse des données a été créé en vue de suivre l'évolution dans ce domaine, dans le but d'essayer d'adapter les techniques existantes aux fins de l'enquête et d'étudier de nouveaux domaines. Le présent document traite de l'un de ces nouveaux domaines.

L'un des outils statistiques de base couramment utilisé actuellement est l'analyse de tableaux de contingence en vue de vérifier l'indépendance de deux caractéristiques ou plus dans une population. L'une des principales hypothèses à la base de la théorie de la distribution est celle de l'indépendance des observations. C'est là l'origine de la distribution multinomiale et, d'une façon asymptotique, du test du chi carré. Si l'hypothèse d'indépendance n'est pas respectée, comme dans une

1. J. Cowan et D.A. Binder, Division des méthodes d'enquêtes institutions et agriculture, Statistique Canada.

enquête par sondage complexe, cette théorie perd sa validité. Plusieurs études ont été menées sur l'analyse de tableaux de contingence là où des éléments de corrélation sont présents entre les observations; mentionnons notamment Cohen [4], Altham [1] et Nathan [6]. Shuster et Downing [7] qui ont établi une statistique de test applicable à tout plan d'échantillonnage.

Ceux-ci se sont demandé dans quelle mesure le plan d'échantillonnage influait sur l'inférence. Est-il possible, par exemple, de supposer que les observations sont indépendantes sans trop déformer l'état réel des choses? La réponse dépendra, bien sûr, du plan d'échantillonnage et de ce que les analystes entendront par « trop ». Une étude est en cours pour examiner de manière empirique l'incidence d'un échantillonnage à deux degrés sur les inférences relatives à l'indépendance de deux caractéristiques d'une population. Précisons que la théorie ainsi conçue est applicable à tout plan d'échantillonnage autopondéré.

2 Distribution des statistiques de test

Désignons la statistique chi carré habituelle de Pearson par P :

$$P = \sum_{i,j} \frac{\left(n\hat{\Pi}_{ij} - n\hat{\Pi}_{i.}\hat{\Pi}_{.j}\right)^2}{n\hat{\Pi}_{i.}\hat{\Pi}_{.j}} ,$$

où les $\hat{\Pi}$ sont les estimations du maximum de vraisemblance (EMV) de $\hat{\Pi}$ dans le cas d'un échantillonnage indépendant et identiquement distribué.

Approximons ensuite P en prenant son développement en série de Taylor autour du point $\{\Pi\colon\ \Pi_{ij} = \Pi_{i.}\Pi_{.j}\}$, où l'hypothèse nulle de l'indépendance se vérifie. On peut aisément montrer que, si $\hat{\Pi}$ est exempt de biais pour Π, les termes constants et linéaires sont nuls. Pour évaluer le terme quadratique, nous considérons :

$$\left.\frac{\delta P}{\delta\hat{\Pi}_{pq}\delta\hat{\Pi}_{rs}}\right|_{\hat{\Pi} = \Pi} = \begin{cases} 2n\left(2 - \dfrac{1}{\Pi_{p.}} - \dfrac{1}{\Pi_{.q}} + \dfrac{1}{\Pi_{pq}}\right) & \text{si } p = r,\ q = s \\ 2n\left(2 - \dfrac{1}{\Pi_{p.}}\right) & \text{si } p = r,\ q \neq s \\ 2n\left(2 - \dfrac{1}{\Pi_{.q}}\right) & \text{si } p \neq r,\ q = s \\ 2n\ (2) & \text{si } p \neq r,\ q \neq s \end{cases} .$$

Soit $A = \frac{1}{2n} \left\{ \frac{\delta P}{\delta \hat{\Pi}_{pq} \delta \hat{\Pi}_{rs}} \middle| \hat{\Pi} = \Pi \right\}$;

dans ce cas, P est approximé par :

$$P \underset{=}{\cdot} n (\hat{\Pi} - \Pi)' A (\hat{\Pi} - \Pi) = Q.$$

Comme $\hat{\Pi}$ est posé comme asymptotiquement normal, la distribution de Q est la somme pondérée de variables aléatoires chi carré centrales indépendantes, chacune à un degré de liberté, les poids (λ_1, λ_2, ...) étant les valeurs propres de la matrice AV (voir Box [3], théorème 2.1). Comme la distribution asymptotique de Q est connue, il est possible de soumettre à une étude empirique l'effet du plan de sondage sur l'inférence en produisant des valeurs de Q et en calculant la proportion de ces valeurs qui dépassent le point critique au niveau α du test du chi carré habituel. Cela pourrait donner une idée de la probabilité qu'une erreur du type 1 soit modifiée par le plan d'échantillonnage.

3 Plans simples à deux degrés

Au moyen de la distribution établie pour le test d'indépendance usuel, nous examinons l'incidence d'un échantillonnage à deux degrés sur les erreurs du type 1 en vérifiant l'indépendance de deux caractéristiques comptant chacune deux catégories. La population est divisée en M unités primaires d'échantillonnage (UPE) desquelles sont tirées m unités. Chaque UPE contient H unités à partir desquelles h unités secondaires d'échantillonnage (USE) sont choisies. Chaque unité échantillonnée est alors classée selon les deux caractéristiques. Ainsi, il est possible de construire un tableau 2 par 2 dans lequel figurent les proportions de la population de départ qui appartiennent respectivement aux quatre catégories, ce qui permet de calculer P. Si les observations avaient été indépendantes, cette statistique aurait asymptotiquement eu une distribution chi carré à un degré de liberté. Nous pouvons vérifier l'hypothèse nulle de l'indépendance des caractéristiques de la population en comparant la valeur observée de la statistique à la valeur en tableau chi carré. Comme les observations ne sont pas indépendantes, le choix du point critique au niveau α provenant de ces tableaux aboutira à un test qui n'est généralement pas au niveau désiré. Comme nous pouvons calculer la distribution asymptotique de P pour des unités m élevées, le niveau réel du test peut être étudié empiriquement en produisant des variables aléatoires avec cette distribution et en calculant la proportion, lesquelles sont supérieures à la valeur critique. Si les observations sont indépendantes, cette proportion correspondra à α en moyenne.

3.1 Méthode

La première étape du calcul de la distribution de P consiste à calculer A, la matrice de la forme quadratique approchant cette statistique. Cela est fait en précisant le paramètre Π, soit les proportions de la population figurant dans chaque catégorie. La prochaine étape consiste à calculer V, qui est la matrice des variances et des covariances de $\sqrt{mh}\ \hat{\Pi}$ pour ce plan de sondage particulier. Nous avons voulu faire de M une variable infinie de manière à tenir compte d'assez près des données de l'Enquête santé Canada, tout en simplifiant les calculs. La façon d'obtenir V est décrite à l'annexe A. Il est nécessaire de spécifier la distribution de ${}_i\Pi = ({}_i\Pi_1, {}_i\Pi_2, {}_i\Pi_3, {}_i\Pi_4)$ sur l'ensemble des UPE dans le calcul de V. Par souci de simplicité, nous avons opté pour la distribution de Dirichlet. L'annexe B indique les propriétés d'une telle distribution.

Nous avons calculé les valeurs propres de AV à l'aide d'un programme conçu par Sparks et Todd [8]; nous avons produit les variables aléatoires normales par un programme venant de Marsaglia et Bray (voir [2]). Pour les valeurs fixes des paramètres, nous avons itéré la valeur approximative de P 10 000 fois, et nous avons enregistré les proportions supérieures aux valeurs critiques aux niveaux 0,10, 0,05, 0,01 et 0,001. Nous avons essayé maintes combinaisons de ces paramètres; les résultats sont présentés à la section suivante.

3.2 Résultats

Dans tous les cas, 3 des 4 valeurs propres sont négligeables et peuvent être tenues pour nulles, de sorte que la distribution de la valeur approximative de P en série de Taylor se ramène à celle d'une variable aléatoire chi carré à un degré de liberté multipliée par la valeur propre la plus élevée (λ_1) de la matrice AV. Cette valeur propre détermine le niveau du test. Si elle est supérieure à 1, l'hypothèse nulle sera rejetée à tort plus de 100 % des fois et, si elle est inférieure à 1, le niveau tombera sous α. Les effets des paramètres sont décrits plus loin.

L'annexe C présente les résultats de certaines des simulations. Ceux-ci sont loin de constituer une liste complète des combinaisons possibles des paramètres, mais cela donne quand même une idée de la façon dont le niveau de signification est modifié par le plan de sondage à deux degrés.

a. Effet de Π (proportions de la population)

Lorsque tous les autres paramètres sont fixes, un changement de proportions de la population dans chacune des quatre catégories se traduit par une variation négligeable de λ_1. Autrement dit, le niveau du test ne dépend pas de Π. Prises isolément, les matrices A, V et AV changent à mesure que Π change, mais la matrice AV aura les mêmes valeurs propres.

b. Effet de h et H (tailles de l'échantillon et de la population dans les UPE)

Si h est fixe, une variation de H détermine une variation limitée de λ_1, mais si H est fixe, une variation de h fait beaucoup varier λ_1 (voir l'annexe C). Le tableau 1 présente les valeurs propres les plus élevées pour diverses valeurs de H et h. Nous pouvons voir qu'une hausse de la taille de la population dans les UPE pour un nombre fixe d'USE dans chaque UPE entraîne une légère hausse du niveau de signification, ce qui est causé par une hausse de la covariance entre les éléments de $\hat{\Pi}$ à mesure qu'augmente H. Le tableau montre également que l'effet de la taille de l'échantillon dans les UPE est de loin plus marqué que celui de la taille de population de ces mêmes UPE et que la fraction de sondage dans les UPE n'est pas informative en soi en raison de l'effet de concentration d'un plus grand nombre d'USE dans les UPE.

Tableau 1

H	h	λ_1
1 000	50	9,13
500	50	9,08
300	50	9,03
100	50	8,75
1 000	40	7,47
800	40	7,46
100	40	7,17
1 000	30	5,81
400	30	5,77
100	30	5,59
1 000	20	4,15
750	20	4,15
500	20	4,13
100	20	4,01
700	15	3,32
400	15	3,30
100	15	3,22
1 000	10	2,49
500	10	2,49
250	10	2,47
100	10	2,42

λ_1 = valeur propre la plus élevée de AV quand $\theta = \sum_{i=1}^{4} \theta_i = 5$

c. Effet de $\theta.$ (paramètre de la distribution des ${}_i\Pi$)

De petites valeurs de $\theta.$ $\left(= \sum_{i=1}^{4} \theta_i\right)$ déterminent de grandes valeurs de λ_1 et, à mesure que $\theta.$ croît à l'infini, λ_1 décroît pour s'établir à $\left(1 - \frac{h}{H}\right)$. Un exemple caractéristique est le cas de H = 300, h = 50 (tableau C4) : $1 - \frac{h}{H} = 0{,}833$.

$\theta.$	λ_1
5	9,030
25	2,727
125	1,226
625	0,915
3 125	0,852
15 625	0,839
78 125	0,837

Ce résultat indique que, plus les caractéristiques sont indépendantes au sein des strates, plus la taille du test est moindre. Fait à noter, si toutes les unités sont échantillonnées dans les strates (h = H), l'indépendance des strates donne une valeur de 0 pour l'approximation de P, comme il se doit. Un autre résultat est que, pour une petite valeur de $\theta.$, une variation de h détermine une grande variation de λ_1, mais que si $\theta.$ est grand, la variation de λ_1 est bien moindre.

4 Orientations futures

Les résultats présentés sont très restrictifs, parce que le modèle est simple. Nous avons opté pour un modèle à deux degrés pour coller le plus près possible à un échantillonnage multinomial. Il reste que, comme le test du chi carré habituel se comporte si mal dans le cas présent, nous devrions nous attendre à ce que les choses empirent dans un cadre plus réaliste.

Un point intéressant est que tous les résultats qui précèdent ont produit une seule valeur propre dominante. En serait-il de même avec des

paramètres plus réalistes (UPE de tailles inégales, échantillons de probabilités inégales, etc.)? Si tel est le cas, nous devrions pouvoir estimer la valeur propre la plus élevée et obtenir une statistique dont la distribution est bien près d'une distribution chi carré. Ajoutons que la distribution de Dirichlet n'était qu'une des nombreuses distributions possibles qui s'offraient à nous.

Annexe A : Calcul de V

Soit u_i = 1 si UPE i est dans l'échantillon $i = 1, \ldots, \infty$
= 0 autrement.

Soit x_{ijk} = 1 si la j^e unité dans la i^e UPE appartient à la catégorie k
= 0 autrement.

$\bar{X}_{i.k}$ est alors la proportion d'unités dans l'UPE i qui appartiennent à la catégorie k;

$\hat{\bar{X}}_{i.k}$ est la proportion d'USE échantillonnées dans l'UPE qui appartiennent à la catégorie k.

Soit $S_{ik\ell} = \dfrac{1}{H-1} \sum_{j=1}^{H} (x_{ijk} - \bar{X}_{i.k})(x_{ij\ell} - \bar{X}_{i.\ell})$.

Nous avons maintenant $y_k = \sum_{i=1}^{\infty} u_i h \hat{\bar{X}}_{i.k}$ = nombre total d'unités échantillonnées qui appartiennent à la catégorie k,

ainsi que $E(y_k) = E_{\mathbf{u}}(E(y_k|u)) = h\, E_{\mathbf{u}}\left(\sum_{i=1}^{\infty} u_i \bar{X}_{i.k}\right) = hm\, E(\bar{X}_{i.k})$.

Nous avons de même

$$\begin{aligned} Cov(y_k, y_\ell) &= E_{\mathbf{u}}(Cov(y_k, y_\ell|\mathbf{u})) + Cov_{\mathbf{u}}[E(y_k|\mathbf{u}), E(y_\ell|\mathbf{u})] \\ &= mh\left(1 - \frac{h}{H}\right) E(S_{ik\ell}) + h^2 m[E(\bar{X}_{i.k}\bar{X}_{i.\ell}) - E(\bar{X}_{i.k})\, E(\bar{X}_{i.\ell})] , \end{aligned}$$

et, comme

$$E(S_{ik\ell}) = \frac{H}{H-1} E\{\delta_{k\ell}\bar{X}_{i.k} - \bar{X}_{i.k}\bar{X}_{i.\ell}\} \begin{pmatrix} \delta_{k\ell} = 1 \text{ si } k = \ell \\ \quad = 0 \text{ autrement} \end{pmatrix},$$

nous avons :

$$\mathrm{Cov}\left(\frac{y_k}{\sqrt{mh}}\ ,\ \frac{y_\ell}{\sqrt{mh}}\right) = \frac{H-h}{H-1}\ E\{\delta_{k\ell}\overline{X}_{i.k} - \overline{X}_{i.k}\overline{X}_{i.\ell}\}$$
$$= h\left[E\left(\overline{X}_{i.k}\overline{X}_{i.\ell}\right) - E\left(\overline{X}_{i.k}\right)\ E\left(\overline{X}_{i.\ell}\right)\right]\ .$$

Annexe B : Distribution de $_i\Pi$

Nous savons que, dans l'UPE i, la somme des proportions dans les diverses catégories est 1 ou $\sum_{j=1}^{4} {}_i\Pi_j = 1$ pour $i = 1, 2, \ldots \infty$. Nous savons aussi que la valeur moyenne est Π_j sur l'ensemble des strates de la proportion dans la catégorie j ou $\lim_{M\to\infty} \frac{1}{M} \sum_{i=1}^{M} {}_i\Pi_j = \Pi_j$ pour $j = 1, 2, 3, 4$. Un moyen simple est de laisser $_i\Pi$ suivre une distribution multidimentionnelle analogue à la distribution bêta et appelée distribution de Dirichlet.

Si le vecteur aléatoire $(y_1, \ldots, y_k)$ épouse une distribution de Dirichlet avec comme paramètres $\theta_1, \ldots, \theta_k$, alors $\sum_{i=1}^{k} y_i = 1;\ y_1 \geq 0,\ i = 1, \ldots, k$, la densité étant donnée par

$$\frac{\Gamma\left(\sum_{i=1}^{k}\theta_i\right)}{\prod_{i=1}^{k}\Gamma\ (\theta_i)}\ \prod_{i=1}^{k} y_i^{\theta_i - 1},$$

$$E(y_i) = \frac{\theta_i}{\Sigma\theta_j}\ ,\ E\left(y_i^2\right) = \frac{\theta_i(\theta_i + 1)}{(\Sigma\theta_j)\ (\Sigma\theta_j + 1)}\ ,$$

$$E(y_i y_k) = \frac{\theta_i\theta_k}{(\Sigma\theta_j)\ (\Sigma\theta_j + 1)}\ .$$

Pour appliquer cette distribution au calcul de V, nous spécifions Π_k, $k = 1, 2, 3, 4$ et faisons correspondre $\frac{\theta_k}{\Sigma\theta_j}$ à Π_k. Cela permet de déterminer $(\theta_1, \theta_2, \theta_3, \theta_4)$ dans les limites d'un multiple constant. Les diverses valeurs de $\theta. = \sum_{i=1}^{4}\theta_i$ dicteront alors les divers degrés d'indépendance des deux caractéristiques à l'intérieur des UPE. À mesure qu'augmente $\theta.$, $\mathrm{Var}\left({}_i\Pi_j\right)$ diminue pour s'établir à zéro pour tous les j, si bien que la distribution des proportions d'unités tombant dans les diverses catégories est identique entre les UPE. Comme la valeur moyenne sur l'ensemble des UPE est Π_k, la distribution à l'intérieur des UPE doit être identique à la proportion de la population et, comme l'indépendance se vérifie dans la population, elle se vérifie nécessairement dans les UPE.

Annexe C : Proportion de fois que la valeur approchée de P dépasse le point critique au niveau de signification donné

C1. Nombre d'unités/UPE = 400

Nombre d'USE/UPE = 30

Valeur propre maximale :	5,77277	2,04550	1,15805	0,97376	0,93662	0,92918	0,92769
θ. est égal à →	5	25	125	625	3 125	15 625	78 125
↓ niveau de signification							
0,10	0,491	0,250	0,130	0,094	0,089	0,088	0,088
0,05	0,412	0,168	0,070	0,046	0,042	0,043	0,042
0,01	0,282	0,071	0,017	0,008	0,007	0,008	0,006
0,001	0,171	0,021	0,002	0,0008	0,0007	0,0006	0,0006

C2. Nombre d'unités/UPE = 400

Nombre d'USE/UPE = 15

Valeur propre maximale :	3,30409	1,50472	1,07630	0,98733	0,96940	0,96581	0,96509
θ. est égal à →	5	25	125	625	3 125	15 625	78 125
↓ niveau de signification							
0,10	0,364	0,175	0,110	0,098	0,093	0,093	0,093
0,05	0,283	0,108	0,057	0,049	0,047	0,044	0,045
0,01	0,157	0,035	0,012	0,009	0,009	0,009	0,008
0,001	0,068	0,007	0,002	0,0009	0,0008	0,0007	0,0009

C3. Nombre d'unités/UPE = 500

Nombre d'USE/UPE = 10

Valeur propre maximale :	2,48497	1,32881	1,05353	0,99637	0,98485	0,98254	0,98208
θ. est égal à →	5	25	125	625	3 125	15 625	78 125
↓ niveau de signification							
0,10	0,300	0,155	0,108	0,097	0,095	0,095	0,094
0,05	0,215	0,088	0,055	0,049	0,048	0,048	0,047
0,01	0,102	0,025	0,012	0,009	0,009	0,009	0,009
0,001	0,037	0,004	0,0012	0,0009	0,0009	0,0009	0,0008

C4. Nombre d'unités/UPE = 300

Nombre d'USE/UPE = 50

Valeur propre maximale :	9,03010	2,72704	1,22631	0,91466	0,85185	0,83927	0,83675
θ. est égal à →	5	25	125	625	3 125	15 625	78 125
↓ niveau de signification							
0,10	0,587	0,320	0,136	0,084	0,073	0,068	0,069
0,05	0,516	0,236	0,077	0,041	0,033	0,030	0,031
0,01	0,395	0,119	0,019	0,007	0,005	0,005	0,005
0,001	0,274	0,046	0,0025	0,0004	0,0004	0,0002	0,0004

C5. Nombre d'unités/UPE = 250

Nombre d'USE/UPE = 10

Valeur propre maximale :	2,46988	1,31140	1,03557	0,97829	0,96674	0,96443	0,96397
θ. est égal à →	5	25	125	625	3 125	15 625	78 125
↓ niveau de signification							
0,10	0,293	0,150	0,104	0,096	0,095	0,092	0,092
0,05	0,212	0,086	0,053	0,047	0,045	0,045	0,045
0,01	0,099	0,025	0,011	0,009	0,008	0,008	0,008
0,001	0,035	0,004	0,0013	0,0010	0,0007	0,0007	0,0008

C6. Nombre d'unités/UPE = 500

Nombre d'USE/UPE = 20

Valeur propre maximale :	4,13494	1,69416	1,11302	0,99234	0,96801	0,96314	0,96217
θ. est égal à →	5	25	125	625	3 125	15 625	78 125
↓ niveau de signification							
0,10	0,415	0,204	0,117	0,098	0,095	0,092	0,092
0,05	0,333	0,131	0,062	0,048	0,045	0,046	0,044
0,01	0,204	0,048	0,014	0,009	0,008	0,008	0,008
0,001	0,105	0,012	0,0020	0,0011	0,0008	0,0008	0,0009

C7. Nombre d'unités/UPE = 750

Nombre d'USE/UPE = 20

Valeur propre maximale :	4,14553	1,70638	1,12563	1,00502	0,98072	0,97585	0,97487
θ. est égal à →	5	25	125	625	3 125	15 625	78 125
↓ niveau de signification							
0,10	0,417	0,211	0,120	0,100	0,093	0,095	0,093
0,05	0,337	0,138	0,064	0,051	0,048	0,047	0,047
0,01	0,207	0,051	0,017	0,010	0,009	0,008	0,008
0,001	0,108	0,012	0,002	0,0012	0,0010	0,0009	0,0007

C8. Nombre d'unités/UPE = 500

Nombre d'USE/UPE = 50

Valeur propre maximale :	9,08484	2,79020	1,29147	0,98023	0,91751	0,90494	0,90243
θ. est égal à →	5	25	125	625	3 125	15 625	78 125
↓ niveau de signification							
0,10	0,582	0,325	0,148	0,095	0,083	0,082	0,083
0,05	0,513	0,244	0,086	0,047	0,040	0,037	0,038
0,01	0,393	0,125	0,023	0,010	0,008	0,007	0,007
0,001	0,275	0,050	0,004	0,0009	0,0007	0,0006	0,0005

Bibliographie

[1] Altham, P.M.E. (1976). "Discrete variable analysis for individuals grouped into families", *Biometrika*, 63, 2, 263-269.

[2] Atkinson, A.C. et Pearce, M.C. (1976). "The computer generation of beta, gamma, and normal random variables", *JRSS*, A, 139, 431-461.

[3] Box, G.E.P. (1954). "Some theorems on quadratic forms applied in the study of analysis of variance problems 1", *AMS*, 25, 2, 290-302.

[4] Cohen, J.E. (1976). "The distribution of the chi-squared statistic under clustered sampling from contingency tables", *JASA* 71, 665-670.

[5] Kish, L. et Frankel, M.R. (1974). "Inference from complex samples", *JRSS(B)*, 36, 1, 1-37.

[6] Nathan, G. (1975). "Tests of independence in contingency tables from stratified proportional samples", *Sankhya (C)*, 37, 77-87.

[7] Shuster, J.J. et Downing, D.J. (1976). "Two-way contingency tables for complex sampling schemes", *Biometrika*, 63, 271-276.

[8] Sparks, D.N. et Todd, A.D. (1973). "Latent roots and vectors of a symmetric matrix", *Applied Statistics*, 22, 2, 260-265.