Identification of Outlying Cells in Multi-way Tables

Jong Cheol Lee* and Chong Sun Hong†

Abstract

An identification method is proposed in order to detect more than one outlying cells in multi-way contingency tables. The iterative proportional fitting method is applied to get expected values of several suspected outlying cells. Since the proposed method uses minimal sufficient statistics under quasi log-linear models, expected counts of outlying cells could be estimated under any hierarchical log-linear models. This method is an extension of the backwards-stepping method of Simonoff (1988) and requires less iteration to identify outlying cells.

keywords : backwards stepping, Bonfferoni's bound, deleted residual, forwards stepping, hierarchical loglinear model, influence, minimal sufficient statistics, outlying cell, quasi-independence.

1 Introduction

When a particular model is tested for fitting contingency table, it is known that some outlying cells might cause significant lack of fit. Various criteria are suggested in order to identify these cells. One kind of these criteria is consisted of residuals, and another is of the difference of goodness-of-fit test statistics. Barnett and Lewis (1994) summarize five criteria among them and recommend three criteria to test suspected cells for the discordancy. They also review exploratory approaches (Mosteller and Parunak 1985) and some other criteria like tetrads which are represented with log odds ratio (Fienberg 1969, Kotze and Hawkins 1984). Haberman (1973) suggests the following adjusted residuals $\tilde{r}_{ij}$ (cf. (1) below) to measure the difference between an observed count x_{ij} and its expected counts $\widehat{m}_{ij}$. Adjusted residuals are plotted on the normal probability plot and could identify extreme cells as outlying cells. Here we note that

*Lecturer, Department of Statistics, Sungkyunkwan University, Seoul 110-745, Korea, jongskku@orgio.net

†Professor, Department of Statistics, Sungkyunkwan University, Seoul 110-745, Korea, cshong@skku.ac.kr

$$\begin{aligned} \tilde{r}_{ij} &= \frac{x_{ij} - \widehat{m}_{ij}}{\sqrt{\widehat{var}(\widehat{m}_{ij})}} \\ &= \frac{x_{ij} - (x_{i+}x_{+j})/N}{\sqrt{x_{i+}x_{+j}(N - x_{i+})(N - x_{+j})/N^3}}, \end{aligned} \tag{1}$$

where $N = \sum_{i=1}^{I}\sum_{j=1}^{J} x_{ij}$, $x_{i+} = \sum_{j=1}^{J} x_{ij}$ and $x_{+j} = \sum_{i=1}^{I} x_{ij}$ are the total count, ith row marginal, and jth column marginal of the table, respectively.

Brown (1974) estimates one omitted cell count, m^*_{ij}, as a missing in $I \times J$ contingency table by

$$m^*_{ij} = \frac{(x_{i+} - x_{ij})(x_{+j} - x_{ij})}{(N - x_{i+} - x_{+j} - x_{ij})}.$$

Then the estimated value is regarded as an observed one which is considered as its expected value under the quasi-independence model (Goodman 1968). The cells influencing too much on goodness-of-fit test statistics are regarded as outlying cells by the criterion of the deleted residual

$$r^*_{ij} = \frac{x_{ij} - m^*_{ij}}{\sqrt{m^*_{ij}}}. \tag{2}$$

When decomposable models are considered for multi-way tables, Upton and Guillen (1995) propose a general formula to estimate one possible outlying cell by extending the deleted residual in (2). However, when the number of minimal sufficient statistics for a certain model increases and there are more than one outlying cell, this formula becomes too complicate to use this formula.

To detect more than one outlying cell, there are two identification methods, among others; one is the forwards-stepping method of Fuchs and Kenett (1980) which tests from the most extreme cell to the least extreme, and the other is the backwards-stepping method of Simonoff (1988) which tests from the least extreme cell to the most extreme. Simonoff notes that the forwards-stepping method could cause the masking effect: the presence of more than one outlying cell causes none to be identified as outlying (a complementary problem is the swamping effect in which the presence of outlying cells, more than one, causes non-outlying cells to be identified outlying). He also notes that the backwards-stepping method using deleted residuals results in the limiting of both the masking and swamping effects. Nonetheless, as the number of the levels of categorical variables increases, two methods tend to take longer time to detect outlying cells. Worse, it is impossible to identify multiple outlying cells for multi-way tables.

In this paper, we propose a method which could identify several outlying cells in multi-way tables. This method uses an iterative proportional fitting (IPF) method to estimate suspected outlying counts. Since this iterative method uses minimal sufficient statistics of certain log-linear model at each iteration steps,

one could estimate suspected outlying cell counts under any hierarchical log-linear models. With this method, more than one omitted cell counts could be estimated at once even in the case of multi-way tables.

By using the deleted residuals, the proposed identification method enables to detect multiple outlying cells for multi-way tables. Since this method is a modified version of the backwards-stepping method, both the masking and the swamping effects might be involved in this method. Nonetheless this method needs smaller number of iterations of calculation than that of the backwards-stepping. The proposed estimation and identification methods are explored with simulation studies.

2 An Iterative Method for Estimating Outlying Cells

To obtain deleted residuals in (2) for two-way table, Brown (1974) treats corresponding cells as missing cells and suggests how to get the expected counts of missing cells. For the identification case of multi-way tables, Upton and Guillen (1995) propose an estimator, say perfect cell value, which can be obtained only for decomposable models. In this section, we develop a method that gives the expected counts of more than one missing cell on multi-way tables for general log-linear models by using the direct extension of the Brown and IPF methods.

For describing outlines of an alternative IPF method in a general situation, let us consider a $I \times J \times K$ table, in which x_{ijk} is the (i,j,k)th cell count. Let T be a set of all cell counts, and let S be a subset of T such that elements in S are supposed to be counts of suspected outlying cells. For three-way tables, it is well-known that the partial association model has no direct solution and that the minimal sufficient statistics of the partial association model are $\{x_{ij+}, x_{i+k}, x_{+jk}\}$. With these statistics, we can construct the following iterative method for estimating cell counts in S.

Iterative Estimation Method

Proc 1. Assign some initial values m_{ijk}^{q} at step $q = 1$ for $(i,j,k) \in S$. For example, $m_{ijk}^{q} = 1$.

Proc 2. Under a given quasi log-linear model, proceed the following fitting steps with minimal sufficient statistics until a desired accuracy for the corresponding cell is attained. For $r = 1, 2, 3, \ldots$,

1. $m_{ijk}^{*q(3r)} = m_{ijk}^{*q(3r-1)} \frac{x_{+jk}^{*q}}{m_{+jk}^{*q(3r-1)}}$,
2. $m_{ijk}^{*q(3r-1)} = m_{ijk}^{*q(3r-2)} \frac{x_{i+k}^{*q}}{m_{i+k}^{*q(3r-2)}}$,
3. $m_{ijk}^{*q(3r-2)} = m_{ijk}^{*q(3r-3)} \frac{x_{ij+}^{*q}}{m_{ij+}^{*q(3r-3)}}$,

where $x_{ij+}^{*q} = x_{ij+} - x_{ijk} + m_{ijk}^{*q(3r-2)}$, $x_{i+k}^{*q} = x_{i+k} - x_{ijk} + m_{ijk}^{*q(3r-1)}$, $x_{+jk}^{*q} = x_{+jk} - x_{ijk} + m_{ijk}^{*q(3r)}$. Note $\widehat{m}_{ijk}^{*q(0)}$ is regards as the initial value.

Proc 3. The value m_{ijk}^{*q} obtained in qth step is regarded as an initial value of the corresponding cell in $(q+1)th$ step. Then repeat iteration steps in <Proc 2> for $q = 1, 2, \ldots$ until a desired accuracy between m_{ijk}^{*q} and $m_{ijk}^{*(q+1)}$ in S is attained.

Then the converged value m_{ijk}^{*q} might be considered as an estimate of (i, j, k) cell in S. Even if any positive values are given as initial values, above iterative method is not affected by any choice of initial values. Positive initial values always make marginal sums positive, so that estimation values of outlying cells in S converge by properties of the IPF method.

Estimated values under the quasi-independence model in two-way tables are obtained by the same way as Brown (1974) used. While Brown's method estimates a single omitted cell value, this proposed method can do for multiple omitted cells in S.

In multi-way tables, perfect cell values of Upton and Guillen (1995) could be used only to decomposable models. However, proposed estimators could be obtained more than one suspected cell and for any log-linear models in multi-way tables as the expected values of the corresponding cells.

3 Identification Method of Multiple Outlying Cells

Simonoff (1988) propose the backwards-stepping method to identify outlying cells from the least extreme cell to the most extreme cell in $I \times J$ contingency tables. To reduce the swapping effects attributed to adjusted residuals of Haberman (1973), he use the deleted residuals in (2) instead of adjusted residuals to detect outlying cells. As the number of minimal sufficient statistics increases, the backwards-stepping method takes longer time to detect outlying cells because of many more number of calculations. Also, this method is restrictive since it could be applied to only two-way tables.

We propose an alternative method to identify multiple outlying cells at once. Moreover this method might have smaller number iterations than that of the backwards-stepping, and can be applied to general log-linear models for multi-way tables.

In this paper, some suspected outlying cells having extreme deleted residuals are put into the set S_q, and the set S_{q+1} is consisted with next considering outlying cells. By comparing S_q with S_{q+1}, outlying cells could be identified. That is, testing hypotheses could be obtained as

$$\begin{array}{ll} H_0 & : \textit{ All elements in } S_{q+1} \textit{ are outlying cells.} \\ H_1 & : \textit{ All elements in } S_q \textit{ are outlying cells.} \end{array} \tag{3}$$

The test statistic for the hypotheses (3) is

$$\Delta G_q^2 = G_{S_{q+1}}^2 - G_{S_q}^2, \tag{4}$$

which is the difference between two generalized likelihood ratio test statistics. One notes that this statistic (4) is identical with Cook's D-statistic to measure the influence of the corresponding cells (Christensen, 1990). Hence the hypotheses (3) could be transformed to

$$\begin{aligned} H_0 &: \textit{The cells which do not belong to } S_{q+1} \textit{ but } S_q \\ &\qquad \textit{are not influential cells.} \\ H_1 &: \textit{The cells which do not belong to } S_{q+1} \textit{ but } S_q \\ &\qquad \textit{are influential cells.} \end{aligned}$$

The test statistic ΔG_q^2 in (4) follows χ^2-distribution with some degrees of freedom which is the difference between the number of elements in two sets S_q and S_{q+1}. When ΔG_q^2 has smaller value, it means that there is no big difference between two goodness-of-fits. Hence elements (cells) in S_{q+1} might be considered as outlying cells. On the other hand, ΔG_q^2 having larger value means that elements in S_q might be regarded as outlying cells. And it could conclude that some elements which do not belong to S_{q+1} but S_q are not outlying cells. In this method, we select as many suspected outlying cells as possible at the initial step in order to reduce the masking effects. Thus, this method tests from the least extreme cell to the most extreme with similar arguments of the backwards-stepping method. Therefore, we knew that the number of initially suspected outlying cells in S_1 is larger than that of S_2 which in turn is larger than that of S_3 ....

Now we propose MOCI method which proceeds the following steps.

MOCI (Multiple Outlying Cells Identification) method

Step 1. Fit an appropriate model to a given contingency table. If it dose not fit well, proceed <Step 2>. Set $q = 1$.

Step 2. Calculate deleted residuals using (2) for all cells. Select some suspected outlying cells that have large values of deleted residuals and put these cells into S_q. Here the values of deleted residuals are compared with a Bonferroni bound ($\Phi^{-1}(1 - \alpha/k)$, where α is a significant level and k is number of cells under consideration) in order to select suspected cells.

Step 3. After estimating the expected counts of suspected cells in S_q by the proposed method in Section 2, recalculate deleted residuals of the cells in S_q. If an absolute value of an deleted residual does not exceed a Bonferroni bound, the corresponding cell is excluded from S_q and construct a new set S_{q+1}.

Step 4. The statistic ΔG_q^2 in (4) could compare two sets S_q and S_{q+1}. If ΔG_q^2 has small value, then the cells in S_{q+1} could be identified as outlying cells. Otherwise, go back to <Step 3> to get deleted residuals of cells in S_{q+1}.

A 5×5 data in <Table 1> is artificially made by Simonoff (1988) for describing the swamping effect, where (1,2), (1,3), and (2,1) cells are set as outlying.

Using proper tables, we would compare the MOCI method with the backwards-stepping method.

Table 1: Artificial Table by Simonoff(1988).

	column				
row	1	2	3	4	5
1	18	41	41	20	21
2	39	20	20	22	22
3	24	20	20	16	18
4	20	20	19	19	19
5	23	19	20	17	20

Expected values, m^*_{ij}, and deleted residuals, r^*_{ij}, in (2) of all cells are calculated under the quasi-independence model and shown in <Table 2>.

Table 2: Estimates of the Omitted Cells

	column				
row	1	2	3	4	5
1	41.9228	23.3728	23.3728	26.1050	28.0473
	(-3.6948)	(3.6461)	(3.6461)	(-1.1949)	(-1.3307)
2	20.4	30.7463	30.7463	20.0331	22.0672
	(4.1181)	(-1.9380)	(-1.9380)	(0.4395)	(-0.0143)
3	20.5556	21.6667	21.6667	16.7435	17.3545
	(0.7597)	(-0.3581)	(-0.3581)	(-0.1817)	(0.1549)
4	22.4314	21.3296	21.8833	15.1554	16.6263
	(-0.5134)	(-0.2879)	(-0.6164)	(0.98756)	(0.5821)
5	21.4413	22.5698	22.0056	16.5288	16.6755
	(0.3366)	(-0.7514)	(-0.4275)	(0.1159)	(0.8141)

* ( ) : the deleted residuals.

One can easily found that the (1,1), (1,2), (1,3), and (2,1) cells have significant discordancy from observed cell counts. The values of deleted residuals in these cells are larger than a Bonferroni bound (2.87 with a significant level 0.05). Hence the four suspected cells are put in S_1. That is, $S_1 = \{(1,1),(1,2),(1,3),(2,1)\}$. In next step, expected cell counts and deleted residuals of cells in S_1 are obtained by the iterative method discussed at <Proc. 2> and <Proc. 3> in Section 2, which are shown at the column of Step 1 in <Table 3>. Note that the estimated count of (1,1) cell is much different from that in <Table 2> and that the absolute value of its deleted residual is smaller than a Bonferroni bound (2.24 with a significant level 0.05). After deleting (1,1) cell in S_1, set $S_2 = \{(1,2),(1,3),(2,1)\}$. Expected cell counts and deleted residuals of cells in S_2 are listed in the Step 2 of <Table 3>. The cell estimates in Step 2 of <Table 3> are not much different from those in Step 1. These deleted residuals in S_2 have larger values than a Bonferroni bound (2.13 with a significant level 0.05). The statistic $\Delta G^2_q = G^2_{S_2} - G^2_{S_2}$ has a value of 1.25 (p-value=0.263 with 1

degree of freedom). Therefore, we might conclude that the (1,2), (1,3) and (2,1) cell in S_2 are outlying cells for <Table 1>.

Note that our method gives the same conclusion as that of the backwards-stepping. However, the MOCI method used only two steps to identify outlying cells and takes much shorter time than that of the backwards-stepping of Simonoff (1988), since Simonoff's method consider 20–30% cells of whole table as suspected outlying cells. The masking effects are reduced much by exploring more than one suspected outlying cells, and influences of corresponding cells are also examined by ΔG_q^2 with the same as Cook's D-statistic.

Table 3: The selection steps.

		Step 1		Step 2	
cell	count	cell estimate	r_{ij}^*	cell estimate	r_{ij}^*
(1,1)	18	24.5981	-1.3304		
(1,2)	41	21.1699	4.3099	19.1567	4.9907
(1,3)	41	21.1699	4.3099	19.1567	4.9907
(2,1)	39	24.7930	2.8532	23.3095	3.2499
$G_{S_q}^2$		$G_{S_1}^2 = 1.34$		$G_{S_2}^2 = 2.59$	
ΔG_q^2		$\Delta G_1^2 = G_{S_2}^2 - G_{S_1}^2 = 2.59 - 1.34 = 1.25$			

4 Simulation Study for 5×5 Tables

Properties of our method are examined and compared with the backwards-stepping of Simonoff (1988) when quasi-independence models are fitted to 5×5 contingency tables. By the simulation studies, behaviors of outlying-cells-identification procedure can be explored with respect to three key attributes.

1. Is the true level of the test close to the significant level α=0.05?
2. How often are cells that are outlying identified as such to get information about the power of the tests?
3. How often are cells that are outlying not identified as such (masked)?

As Simonoff (1988) did, tables of size 5×5 were examined in this simulation. These tables were generated using the International Mathematical and Statistical Libraries (IMSL) routine GGMTN (IMSL 1982); there were 1,600 replications for each simulation run to compare the MOCI with the backwards-stepping. The test based on deleted residuals uses a Gaussian reference, whereas a difference in G^2 is compared with χ^2 distributions.

Simulations exploring various alternatives were examined, which gives the power ($\beta_1 = \Pr$(detecting any cell outlying)), the probability of identifying outlying cells and only those cells correct (β_2), the probability of at least one cell identified correct (β_3), the average number of cells identified correct (N_c), and

the average number of cells identified incorrect (N_i). The alternatives are defined by

$$p^*_{ij} = \begin{cases} p_{ij}(1+\Delta_{ij}/\sqrt{N}), & (i,j) \in S, \\ p_{ij}\big(1-\sum_S p^*_{ij}\big)/\sum_{T|S} p_{ij}, & (i,j) \in T|S, \end{cases} \tag{5}$$

where p_{ij} is the null probabilities, S is the set of outlying cells, and $T|S$ is the set of non-outlying cells. The null probabilities were uniform for a 5×5 table. The table refers to four alternatives with $N = 500$ and three outlying cells : (i) $\Delta_{11} = \Delta_{12} = \Delta_{13} = 30$, (ii) $\Delta_{11} = 20, \Delta_{12} = 30, \Delta_{13} = 40$, (iii) $\Delta_{11} = -20, \Delta_{12} = 20, \Delta_{13} = 40$, (iv) $\Delta_{11} = 30, \Delta_{22} = 30, \Delta_{33} = 30$.

Table 4: Comparison of detection procedures under three multiple outlying alternatives to independence for 5×5 table.

	β_1	β_2	β_3	N_c	N_i	N_r
(i) $\Delta_{11} = \Delta_{12} = \Delta_{13} = 30$						
Backwards-stepping	0.806	0.206	0.389	0.908	0.795	115
MOCI	0.994	0.000	0.318	0.426	1.773	33
(ii) $\Delta_{11} = 20$, $\Delta_{12} = 30$, $\Delta_{13} = 40$						
Backwards-stepping	0.851	0.212	0.627	1.288	0.470	115
MOCI	0.998	0.002	0.577	0.787	1.660	33
(iii) $\Delta_{11} = -20$, $\Delta_{12} = 20$, $\Delta_{13} = 40$						
Backwards-stepping	1.000	0.281	0.999	2.260	0.579	115
MOCI	1.000	0.389	1.000	2.403	0.555	33
(iv) $\Delta_{11} = 30$, $\Delta_{22} = 30$, $\Delta_{33} = 30$						
Backwards-stepping	1.000	0.655	1.000	2.695	0.069	115
MOCI	1.000	0.768	1.000	2.956	0.263	33

* β_1 is the power. β_2 is the probability of identifying outlying cells and only those cells correct. β_3 is the probability of at least one cell identified correct. N_c is the average number of cells identified correct. N_i is the average number of cells identified incorrect. N_r is the average iteration number of estimation of the deleted residual.

Since N_c is 0.908 and 1.288 in the backwards-stepping and 0.426 and 0.787 in the MOCI for (i) and (ii) cases of <Table 4>, respectively, the MOCI method has smaller possibilities identifying correct outlying cells than those of the backwards-stepping. However, N_c in the MOCI has larger value in (iii) and (iv). Since minimal sufficient statistics play very important role in the MOCI, it is much more susceptible to masking effects than that of the backwards-stepping when outlying cells exist on one row or one column. When minimal sufficient statistics are much less affected in both (iii) and (iv) cases, the MOCI is much less susceptible to masking effects than the backwards-stepping, that is, β_2 and N_c for the MOCI have larger values than those of the backwards-stepping. Moreover, the backwards-stepping needs many number of iterations because it should calculate deleted residuals of almost all cells at each step, whereas the

MOCI requires much less number of calculations since only the deleted residuals of corresponding cells are obtained. The average iteration number of estimation for the deleted residual (N_r) for the backwards-stepping is about 115 (=25+24+23+22+21) , but that for the MOCI is only 33 (=25+5+3). If the number of categorical level increases, we could conclude that the MOCI takes much less computing time than that of the backwards-stepping.

5 Simulation Study in the Multi-Way Tables

In this section, simulation for detecting outlying cells in multi-way tables are examined. Consider the case of $3 \times 3 \times 3$ tables to explore the efficiency of detecting outlying cells in general multi-way tables. Let $[AB][C]$ be an appropriate model to fit, and then there exist direct solutions to the model. The expected count m_{ijk} could be obtained by using the following log-linear model

$$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)}.$$

Since (i, j, k) cell probability is defined as $p_{ijk} = m_{ij+}m_{++k}/N^2$, we would obtain the random sample of size $3 \times 3 \times 3$ table. The alternatives in the simulations are as in (5). With similar arguments as those of two-way tables, we could have examined the power (β_1), the probability of identifying outlying cells and only those cells correct (β_2), the probability of at least one cell identified correct (β_3), the average number of cells identified correct (N_c), and the average number of cells identified incorrect (N_i). For model $[AB][C]$, simulation results for identifying outlying cells are shown in <Table 5>.

Table 5: Simulation Results for Identifying the Outlying Cells in $[AB][C]$.

	β_1	β_2	β_3	N_c	N_i	N_r
$\Delta_{111} = 30$	1.000	0.059	1.000	1.000	2.583	34
$\Delta_{111} = 30, \Delta_{332} = 30$	1.000	0.040	1.000	1.636	2.482	34
$\Delta_{111} = 30, \Delta_{332} = 30, \Delta_{223} = 30$	1.000	0.054	1.000	2.485	2.345	36

Table 6: Simulation Results for Identifying the Outlying Cells in $[AB][AC][BC]$.

	β_1	β_2	β_3	N_c	N_i	N_r
$\Delta_{111} = 30$	0.978	0.101	0.853	0.853	2.439	34
$\Delta_{111} = 30, \Delta_{332} = 30$	0.986	0.057	0.917	1.410	2.890	35

In <Table 6>, consider the following partial association model $[AB][AC][BC]$ which does not have direct solutions.

$$\begin{aligned} \log m_{ijk} &= u + u_{1(i)} + u_{2(j)} + u_{3(k)} \\ &\quad + u_{12(ij)} + u_{13(ik)} + u_{23(jk)}. \end{aligned}$$

<Table 5> and <Table 6> show that N_c is smaller and N_i is larger than the number of correct outlying cells, which means that the MOCI is much susceptible to swamping. These results that would tell that the model $[AB][AC][BC]$ also involves the swamping effect such as $[AB][C]$ model does. In particular, very small value of β_2 means that it is not easy to identify outlying cells and only those cells correct. We need to remember that the MOCI can identify outlying cells with much smaller number of calculations in multi-way tables.

6 Conclusions

When some model is fitted for a contingency table, some outlying cells would cause significant lack of fit for the model. It is well-known that these cells could be detected by using residuals and goodness-of-fit test statistics. Multiple outlying cells might be identified via the forwards-stepping or the backwards-stepping methods, but these can be applied only to two-way tables and may require much longer computation time.

This paper propose the MOCI method which detects several outlying cells simultaneously, and takes much less computing time. Since the MOCI method can be extended to multi-way tables, it could be utilized to various log-linear models for multi-way tables such as linear-by-linear model, row and column effect model, etc. We finally note that our MOCI method can experience some difficulty in identifing outlying cells when there exists several outlying cells in one row or column.

References

[1] Barnett, V., and Lewis, T. 1994. *Outliers in statistical data*, 3rd, John Wiley & Sons.

[2] Brown, M. L. 1974. "Identification of the sources of significance in two-way contingency tables," *Applied Statistics*, Vol. 23, 405-413.

[3] Christensen, R. 1990, *Log-linear models*, New York : John Wiley & Sons.

[4] Fienberg, S. E. 1969. "Preliminary Graphical analysis and quasi-independence for two- way contingency table," *Applied Statistics*, Vol. 18, 153-168.

[5] Fuchs, C., and Kenett, R. 1980. "A test for outlying cells in the multinomial distribution and two–way contingency tables," *Journal of the American Statistical Association*, Vol. 75, 395-398.

[6] Goodman, L. A. 1968. "The analysis of cross-classified data: independence, quasi-independence, and interaction in contingency tables with or without missing cells," *Journal of the American Statistical Association*, Vol. 63, 1091-1131.

[7] Haberman, S. J. 1973. "The analysis of residuals in cross-classified tables," *Biometrics*, Vol. 29, 205-220.

[8] International Mathematical and Statistical Libraries 1982. Subroutine GGMTN, Houston.

[9] Kotze, T. J. W., and Hawkins, D. M. 1984. "The identification of outliers in two-way contingency tables using 2×2 subtables," *Applied Statistics*, Vol. 33, 215-223.

[10] Mosteller, F., and Parunak, A. 1985. "Identifying extreme cells in a sizable contingency table : probabilistic and exploratory approaches," *In Exploring Data Tables, Trends and Shapes*, Wiley, 189-224.

[11] Simonoff, J. S. 1988. "Detecting outlying cells in two-way contingency tables via backwards-stepping," *Technometrics*, Vol. 30, No. 3. 339-345 (with Correction, Vol. 32, No. 1, 115).

[12] Upton, G. J. G., and Guillen, M. 1995. "Perfect cells, direct models and contingency table outliers," *Communications in Statistics, Part A - Theory and Methods*, Vol. 24, 1843-1862.