Generalized Mantel-Haenszel Procedures for 2 × J Tables

Takashi Yanagawa, 1 Yoshinori Fujii, 2 and Joe Mastuoka 3

1Department of Mathematics, Kyushu University, Fukuoka, Japan; 2Department of Mathematics, Faculty of Education, Miyazaki University, Miyazaki, Japan; 3Fujisawa Pharmaceutical Industry Co. Ltd., Osaka, Japan

Generalization of Mantel-Haenszel procedure for 2 × J (J > 2) tables is reviewed. Included are generalized Mantel-Haenszel tests, estimators for a common odds ratio, and generalized Breslow-Day test for the homogeneity of odds ratios across the strata.—Environ Health Perspect 102(Suppl 8): 57–60 (1994)

Key words: contingency tables, cause-effect association, epidemiology, odds ratio, statistical test, statistical estimation

Introduction

The Mantel-Haenszel procedure (1) is a set of the statistical methods most frequently employed in the analysis of epidemiologic data. It consists of estimation and testing of the odds ratio, presuming the homogeneity of the odds ratios across strata. A method for testing the homogeneity of the odds ratios has been developed by Breslow and Day (2) and added to the procedure [for example, SAS version 6.1 (3)]. The procedure is solely for 2 × 2 tables.

Logistic regression analysis, widely used in the analysis of epidemiologic data, employs the unconditional method of maximum likelihood. Occasionally the conditional method of maximum likelihood also is employed in the analysis, in particular when the sample size is small. Compared with these methods, the Mantel-Haenszel procedure has the following characteristics: while the results by logistic regression analysis are sensitive to the mathematical model employed, the Mantel-Haenszel procedure is free from the models; more precisely, the test statistic is nonparametric and the estimator is a moment estimate (σ). Furthermore, the Mantel-Haenszel estimator is dually consistent, i.e., consistent when the number of tables is fixed and the sample size in each stratum is large (5); it is also consistent when the sample size in each table is fixed and the number of tables is large (6). While the unconditional maximum likelihood estimator (MLE) is consistent in the former situation, it is not consistent in the latter (6–8). Thus, the Mantel-Haenszel estimator is more stable for sparse tables than the unconditional MLE. It has been shown that when subject responses are correlated within tables, the Mantel-Haenszel estimator is consistent, while the conditional MLE is not consistent (9).

Birch (10), Landis, Heyman, and Koch (11), Mantel and Byar (12), and Yanagawa (13) extended the Mantel-Haenszel test for 2 × J tables. Yanagawa and Fujii (14) generalized the Breslow-Day test for 2 × J tables. Liang (15), Greenland (16), and Yanagawa and Fujii (14) extended the Mantel-Haenszel estimator for 2 × J tables. The purpose of the present article is to review the generalization of the Mantel-Haenszel procedure for 2 × J tables.

Notation

We consider K 2 × J tables for which the cell frequencies and cell probabilities are shown in Table 1. Our primary interest is the relationship between the two classifications represented by the second (disease) and the third (exposure) indices; the first index denotes nuisance variables on which we are stratifying. The j'th odds ratio taking the j'th column as a base in the k'th stratum will be denoted by:

ψjk = Pk1j/Pk2j

(j = 2,..., J; k = 1, 2,..., K). The Mantel-Haenszel type testing and estimation presume the homogeneity of the odds ratios; more specifically

ψjk = ψj+ = ... = ψjJ (= ψjk, say), [1]

and focus on the common odds ratios, ψj+}. It is easy to see that the odds ratios satisfy the following property:

ψjk = ψj+ = 1, ψj+ = 1/ψj

ψjk = ψj+ = ψj

for all j, j', j'' = 1, 2,..., J. In particular, putting ψj+ = ψj+ we have ψjk = ψjk/ψj.

Thus, under Equation 2, ψjk,j' = 1, 2,..., J, play the basic role in the inference below.

The property (Equation 2) is called the invariance of the odds ratios with respect to the selection of the comparison group.

Testing Exposure and Disease Association

We note that no exposure and disease association is expressed by ψj,j' = 1 for any j, j' = 1, 2,..., J, which is equivalent to ψj = 1, j = 2, 3,..., J.

Table 1. The cell frequencies and cell probabilities in the kth table.

| Cell frequencies | Cell probabilities |
|------------------|--------------------|
| x11, x12, ..., x1j, ..., x1J | Pk11, Pk12, ..., Pk1j, ..., Pk1J |
| x21, x22, ..., x2j, ..., x2J | Pk21, Pk22, ..., Pk2j, ..., Pk2J |
| x1, x2, ..., xj, ..., xJ | P̂k1, P̂k2, ..., P̂kj, ..., P̂kJ |

This paper was presented at the 4th Japan-US Biostatistics Conference on the Study of Human Cancer held 9–11 November 1992 in Tokyo, Japan.

Address correspondence to T. Yanagawa, Department of Mathematics, Kyushu University 33, Fukuoka 812, Japan. Telephone 011 81 92 801 3751. Fax 011 81 92 632 2737.
The Case of Nonordinal Exposure

For testing the null hypothesis $H_0: \psi_j = 1$, $j = 1, 2, \ldots, J$ against $H_1: H_0$ is not true, we use the test statistic given by Yanagawa (13) which follows chi-square distribution with $(K-1)(J-1)$ under $H_0$ asymptotically.

The Case of Ordinal Exposure

We suppose that the values $C_{jk1}, C_{jk2}, \ldots, C_{jkj}$ ($C_{jk1}, C_{jk2}, \ldots, C_{jkj}$) may be given to exposure categories. Putting $\psi_{jk} = \beta(C_{jk} - C_{k-})$, we formulate the problem by testing the null hypothesis $H_0: \beta = 0$ against $H_1: \beta \neq 0$ or $\beta > 0$. We have the one degree of freedom chi-square test given by Landis, Heyman, and Koch (11, 13). In particular, if $C_{jk} = C_j$ for all $k$, then the test reduces to the one proposed by Birch (10) and Mantel-Byar (12).

Estimating Exposure and Disease Relationship

A Projection Method

The Mantel-Haenszel estimator of the common odds ratio from $2 \times 2$ tables that are made by the $j$ and $j'$ columns in Table 1 is denoted by:

$$a_{jk}^{(j)} = R_{yj} / R_{yj'},$$

where

$$R_{yj} = \sum_{k=1}^{K} x_{yjk} x_{yjk'}, x_k.$$  

We call the $\{R_{yj}\}$ the naive Mantel-Haenszel estimator. They lack the invariance property with respect to the selection of a base when $j > 2$; namely,

$$a_{jk}^{(j)} a_{jk'}^{(j')} \neq a_{jk}^{(j)}.$$

To modify these estimators to have the invariance property, we introduce a projection method. Put

$$\theta_j = 0$$

$$\theta_j = \log \psi_{jk}, j = 2, 3, \ldots, J$$

$$b_{jk}^{(j)} = \log a_{jk}^{(j)}$$

and

$$Q = \sum_{j=2}^{J} W_{yj} (b_{jk}^{(j)} - (\theta_j - \theta_j)),$$

where $W_{yj}$ is an appropriate weight such that $W_{yj} = W_{yj'}$ and is assumed independent of $[\theta_j]$. The estimates of $[\theta_j]$ are given by minimizing $Q$ with respect to $\theta_j$.

When these estimates are obtained, the estimate of the odds ratio $\psi_{jk}^{(j)}$ is given by

$$\psi_{jk}^{(j)} = \exp \left[ \theta_j - \theta_j \right],$$

which is easily seen to be an invariant estimator with respect to the selection of a base. We call this estimator a projection method estimator. A formula for the estimated variance of the projection method estimator is given by Yanagawa and Fujii (17).

Example 1. Putting $W_{yj} = 1$ for all $j$ and $j'$, we have

$$\hat{\theta}_j = \frac{1}{J} \sum_{k=1}^{K} \left( b_{jk}^{(j)} - b_{k1}^{(j)} \right),$$

which is the generalized Mantel-Haenszel estimator of Greenland (16) and Yanagawa and Fujii (14).

Example 2. Putting $\theta_j = \beta C_{jk} j = 1, 2, \ldots, J$, where $C_{jk}$'s are the values of the ordinal categories, the estimator of $\beta$ is given by:

$$\hat{\beta} = \left[ \sum_{j=2}^{J} \rho_{yj} (C_{jk} - C_{jk'}) \right]^{-1} \left[ \sum_{j=2}^{J} \rho_{yj} (C_{jk} - C_{jk'}) \right].$$

Liang's Estimating Function

Liang (15) generalized the Mantel-Haenszel estimator by introducing an estimating function. In the simplest case, his estimating function may be represented by:

$$Q = \sum_{j=2}^{J} \frac{1}{W_{yj}} \left[ R_{yj} \psi_{jk} - R_{yj'} \psi_{jk'} \right] = 0.$$

The estimate of $\psi = (\psi_2, \ldots, \psi_J)$ is obtained from this equation by means of an iterative method. We may show that the projection method estimator with the weight $\rho_{yj} = R_{yj} / R_{yj'} (R_{yj} + R_{yj'})$ is asymptotically equivalent to the Liang estimator. Note that, since $a_{jk}^{(j)}$ is dually consistent, the asymptotic equivalence referred to includes both where the number of tables is fixed and the sample size in each stratum is large, and where the sample size in each table is fixed and the number of tables is large.

Testing the Homogeneity of the Odds Ratios

The Mantel-Haenszel type testing and estimation procedures in the preceding sections presume the homogeneity of the odds ratios presented in Equation 1. We generalize the Breslow-Day test for testing

$$H_0: \psi_{jk} = \psi_{j}^{(j)} = \psi_{j}^{(j)} = \psi_{j}^{(j)} = \ldots, \psi_{j}^{(j)} = \psi_{j}^{(j)},$$

against

$$H_1: H_0 \text{ is not true}.$$

Putting

$$X'_i = (x_{i2}^{(1)}, \ldots, x_{iJ}^{(1)}),$$

$$\psi' = (\psi_2, \ldots, \psi_J),$$

$$E_{i}(\psi'), V_{i}(\psi'),$$

the asymptotic conditional mean vector and covariance matrix of $X'_i$, conditioned on all marginals $x_{i1}^{(1)}, x_{iK}^{(1)}, (i=1, 2, j=1, \ldots, J, k=1, \ldots, K),$ the Breslow-Day test is generalized to be

$$S(\tilde{\psi}_{MH}) = \frac{1}{K} \sum_{k=1}^{K} \left[ x_k - E_i(\tilde{\psi}_{MH}) \right]$$

$$V_{i}(\tilde{\psi}_{MH})^{-1} \left[ x_k - E_i(\tilde{\psi}_{MH}) \right],$$

where $\tilde{\psi}_{MH}$ is the generalized Mantel-Haenszel estimator of Greenland (16) and Yanagawa and Fujii (14). Yanagawa and Fujii (14) show that this statistic does not follow chi-square distribution under $H_0$ asymptotically and that it affords anticonservative results. A method of modification of this statistic is given by Yanagawa and Fujii (14) so that it follows chi-square distribution with $(K-1)(J-1)$ d.f. asymptotically under $H_0$, together with the method of modification of the generalized Mantel-Haenszel estimator so that it tends to be asymptotically efficient. The algorithm for the computation is represented as follows:

Step 1. Obtain the generalized Mantel-Haenszel estimate $\tilde{\psi}_{MH}$.

Step 2. Obtain the mean vector $(E_k)$ and covariance matrix $(V_k)$ of $X_k$; this is undertaken by iterative proportional fitting (IPF) in each stratum by putting $q_{j1} = 1, q_{j2} = \psi_{MH}(j = 1, 2, \ldots, J)$ as initial values for the cell frequencies. Let $m_{kij}$ be the converged frequency of the $(i, j)$ cell in the $k$th stratum. Then compute $V_k = E_k + D(k), E(k)$ is the unit matrix of order $J-1; D(r)$ is the diagonal matrix with elements $r_1, r_2, \ldots, r_j$; and $r_i = (1/m_{kij}) + (1/m_{kij})$. The estimate of the $j$th element of $E_k$ is $m_{kij}$ and the estimate of $V_k$ is the inverse of $V_k^{-1}$.
Step 3. Compute the chi-square statistic

\[ S(\Psi_{MH}) = \sum_{i=1}^{K} \sum_{j=1}^{J} \left( \frac{(X_{ij} - m_{ij})^2}{m_{ij}} + \frac{(X_{kj} - m_{kj})^2}{m_{kj}} \right) \]

Step 4. Compute the efficient estimate of the common odds ratio

\[ \hat{\Psi} = \hat{\Psi}_{MH} + D(\hat{\Psi}_{MH}) \left( \sum_{i=1}^{K} V_i \right)^{-1} \]

where \( M_i = (m_{k2}, \ldots, m_{k2}) \).

Step 5. Compute the chi-square statistic

\[ T(\Psi_{MH}) = S(\Psi_{MH}) - \left( \sum_{i=1}^{K} X_i - M_i \right)^2 \left( \sum_{i=1}^{K} V_i \right)^{-1} \left( \sum_{i=1}^{K} X_i - M_i \right) \]

Computer Program

Using SAS/IML (3) we made a computer program for the Mantel-Haenszel procedure for 2×J tables described in this article. We wish to publish it elsewhere.

An Example

As an illustration of the generalized Mantel-Haenszel procedure, we use part of a data set from the case-control study of esophageal cancer given in Breslow and Day (2). Suppose that our primary concern is the association of alcohol consumption and esophageal cancer among persons less than 55 years of age. Epidemiologists might prefer stratifying on the confounding variables (in this case age and tobacco consumption) and then, if the odds ratios are homogeneous throughout the strata, summarizing the information in each table by estimating and testing the common odds ratios. The method is intuitive and easy to comprehend, but from the point of view of analysis we have to deal with an abundance of small entries and empty cells. The data are summarized in eight 2×4 tables, shown in Table 2. The response categories are alcohol consumption of 0 to 39, 40 to 79, 80 to 119, and 120+ g/day. The results of analysis by the generalized Mantel-Haenszel procedure and those by a log-linear model are presented in Table 3. Alcohol consumption of 0 to 9 g/day has been taken as a base in the table. The tendency of the unconditional MLE from the log-linear model towards inflated values with sparse data is evident for the 120+ g/day category; in this case the generalized Mantel-Haenszel estimate is 121.80, and it is improved to be 194.23. The generalized Mantel-Haenszel estimator has little influence on sparse data (16). The improvement is essential for the homogeneity test to follow a chi-square distribution. The value of the corrected chi-square homogeneity test is 24.66; whereas \( G^2 = 26.19 \) and \( X^2 = 27.73 \). Although the difference is not as great as in the case of estimation, we can still see the impact of the sparse data on \( G^2 \), the unconditional likelihood ratio and on \( X^2 \), the Pearson chi-square.

| Age | Tobacco | 0-39 | 40-79 | 80-119 | 120+ | Total |
|-----|---------|------|-------|--------|------|-------|
| 25-34 | 10-19 | Case | 0 | 0 | 1 | 1 |
| 35-44 | 0-9 | Case | 10 | 7 | 1 | 0 | 18 |
| 35-44 | 10-19 | Case | 1 | 3 | 0 | 0 | 4 |
| 45-54 | 0-9 | Case | 1 | 6 | 3 | 4 | 14 |
| 45-54 | 10-19 | Case | 0 | 4 | 6 | 3 | 13 |
| 45-54 | 20-29 | Case | 18 | 17 | 8 | 1 | 44 |
| 45-54 | 30+ | Case | 0 | 5 | 1 | 2 | 8 |
|       |       | Control | 10 | 10 | 4 | 1 | 25 |
|       |       | Control | 0 | 5 | 2 | 4 | 11 |

Table 3. Results of the analysis.

| Alcohol consumption, g/day | 0-39 | 40-79 | 80-119 | 120+ |
|---------------------------|------|-------|--------|------|
| The estimated odds ratio   |      |       |        |      |
| The generalized Mantel-Haenszela | 10.11 | 11.45 | 121.80 |
| The improved estimate      | 13.29 | 14.87 | 194.23 |
| The unconditional MLE      | 14.22 | 16.08 | 235.12 |
| The homogeneity test chi-square (df=21) |      |       |        |      |
| Uncorrected                | 25.30 |       |        |      |
| Corrected                  | 24.66 |       |        |      |
| Likelihood ratio chi-square| 26.19 |       |        |      |
| Pearson chi-square         | 27.73 |       |        |      |

*From Breslow and Day (2).

Table 3. Results of the analysis.

| Alcohol consumption, g/day | 0-39 | 40-79 | 80-119 | 120+ |
|---------------------------|------|-------|--------|------|
| The estimated odds ratio   |      |       |        |      |
| The generalized Mantel-Haenszela | 10.11 | 11.45 | 121.80 |
| The improved estimate      | 13.29 | 14.87 | 194.23 |
| The unconditional MLE      | 14.22 | 16.08 | 235.12 |
| The homogeneity test chi-square (df=21) |      |       |        |      |
| Uncorrected                | 25.30 |       |        |      |
| Corrected                  | 24.66 |       |        |      |
| Likelihood ratio chi-square| 26.19 |       |        |      |
| Pearson chi-square         | 27.73 |       |        |      |

*Weight W=1

REFERENCES

1. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 22:719-748; 1959.
2. Breslow NE, Day NE. Statistical Methods in Cancer Research, vol 1, The Analysis of Case-Control Studies. Lyon: International Agency for Research on Cancer, 1980.
3. SAS. SAS User’s Guide: Statistics, Version 5. Cary, NC:SAS Institute Inc., 1985.
4. Yanagimoto T. Combining moment estimates of a parameter common through strata. J Stat Planning and Inference 25:187-198 (1990).
5. Gart JJ. On the combination of the relative risks. Biometrics 18:594-600 (1962).
6. Breslow NE. Odds ratio estimators when the data are sparse.
7. Andersen EB. Conditional Inference and Models for Measuring. Copenhagen: Mentalhygienisk Forlag, 1973.
8. Pike MC, Hill AO, Smith PG. Bias and efficiency in logistic analyses of stratified case-control studies. Int J Epidemiology 9:89–95 (1980).
9. Liang K-Y. Odds ratio inference with dependent data. Biometrika 72:678–682 (1985).
10. Birch MW. The detection of partial association I: the general case. J Roy Stat Soc B27:111–124 (1965).
11. Landis JR, Heman ER, Koch GG. Average partial association in three way contingency tables: a review and discussion of alternative tests. Int Stat Rev 46:237–354 (1978).
12. Mantel N, Byar DP. Marginal homogeneity, symmetry and independence. Comm Statist A7:9953–976 (1978).