Tests of Independence with Incomplete Contingency Tables using Likelihood Functions

Shin-Soo Kang [1], Kenneth J. Koehler [2], and Michael D. Larsen [2]
Department of Information and Statistics, KwanDong University, Kangwon-Do, 210-701, South Korea[1]
Department of Statistics, Iowa State University, Ames, Iowa 50011-1210, U.S.A.[2], larsen@iastate.edu

Abstract

Maximum likelihood estimators (MLEs) of parameters can be obtained from the log-likelihood function for cell probabilities of two-way incomplete contingency tables. Constraints on probabilities can be incorporated through Lagrangian multipliers. Variances of the MLEs can be derived from the matrix of second derivatives of the loglikelihood with respect to cell probabilities and the Lagrangian multipliers. In this paper, Wald and likelihood ratio tests of independence are examined. Simulation results under the assumption that data are missing at random reveal that maximum likelihood esimtation (MLE) produces more efficient estimates of population proportions than either multiple imputation (MI) based on data augmentation or complete case (CC) analysis. Neither MLE nor MI leads to an improvement over CC analysis with respect to the power of tests for independence. Thus, the partially classified marginal information increases the precision of estimates of proportions, but is not helpful for judging independence.

Keywords: Complete Case Analysis; Lagrange Multipliers; Maximum Likelihood Estimation; Multiple Imputation; Wald Statistic.

1. Introduction

In the analysis of contingency tables, it may happen that some observations are not fully cross-classified. One simple approach, known as complete-case (CC) analysis, discards the missing data by restricting analysis to only fully classified counts in an incomplete contingency table.

Chen and Fienberg (1974) used an iterative procedure for computing maximum likelihood estimates and developed Pearson and likelihood ratio tests of independence for two-way tables for which either the row classification or the column classification could be missing for some cases. As in Chen and Fienberg (1974), Hocking and Oxspring (1974) consider three independent multinomial distributions corresponding to the set of fully cross-classified counts and the two sets of partially classified counts, where either the row classification or the column classification is missing.

An alternative approach involves constructing a complete table, in which all cases are completely classified, by imputing information for the missing row or column classification. Multiple imputation, proposed by Rubin (1978; see also Rubin 1987, 1996), provides a way to take advantage of commonly used tests of independence for completely classified tables. Li et al. (1991) proposed a Wald test statistic and Meng and Rubin (1992) proposed likelihood ratio tests with F reference distributions. In addition to such tests of independence, one can estimate joint probabilities and their standard errors using sets of, say, five imputed two-way contingency tables obtained from data augmentation.

West et al. (2002) proposed a method using an EM algorithm (Dempster, Laird, and Rubin 1977) and a Bayesian prior distribution on unknown proportions to analyze longitudinal studies with repeated measures. They generate all possible sets of values for those that are missing, form a set of possible complete data sets, and then weight each data set according to clearly defined assumptions. They apply an appropriate statistical test procedure to each data set, combining the results to give an overall indication of significance.

Instead of imputing information for missing classifications, maximum likelihood estimates of population proportions can be obtained from the observed information, including both the completely and partially classified cases. Little (1982) used a simple EM algorithm to estimate cell probabilities. Lipsitz et al. (1998) show how to use generalized linear model software to evaluate maximum likelihood estimates of cell probabilities using the connection between the multinomial and Poisson likelihoods.

Hot deck imputation methods select observed cases to 'donate' values to fill-in missing entries. The Fully Efficient Fractional Imputation (FEFI) method described by Kang, Koehler, and Larsen (2006) uses all feasible observed cases as donors thereby reducing variance due to random imputation. It yields consistent estimators of cell probabilities if the missing mechanism is missing completely at random (MCAR; Rubin 1976). It is not necessarily consistent if data are simply missing at random (MAR).

Maximum likelihood estimates of population proportions are obtained from the partial log-likelihood function for the cell probabilities of two way incomplete contingency tables. Chen and Fienberg (1974) used this log likelihood in their work. This MLE method does work under MAR. Constraints such as the fact that cell probabilities sum to one can be incorporated with Lagrangian multipliers, as proposed by Aitchison and Silvey (1958). Variances of MLE estimators of population proportions can be derived from the matrix of second derivatives of the loglikelihood with respect to cell probabilities. Wald and likelihood ratio tests of independence can be based on the likelihood function evaluated at the MLEs.

Section 2 presents notation and the likelihood function. Section 3 derives MLEs and their variances. Section 4 discusses tests of independence. The performance of tests of independence provided by the complete-case analysis, multiple imputation, and maximum likelihood approaches are examined through Monte Carlo simulation studies in section 5 considering both type I error level and power. Section 6 contains a summary and discussion.

2. Notation and Likelihood Function

Consider an $I \times J$ contingency table where the row factor X_1 has I categories and the column factor X_2 has J categories. Assume simple random sampling with replacement. In a complete table, where the row and column categories are observed for every case in the sample, the counts have a multinomial distribution with sample size N and probability vector θ. Let n_{ij} denote the count for the (i, j) cell, and let θ_{ij}, an element of θ, denote the population proportion for the (i, j) cell.

When information on either the row or column classification is missing, we can construct a table of counts for the completely classified cases where x_{ij} denotes the number of cases observed in the (i, j) cell. We can also construct one-way tables of counts for partially classified cases. Let x_{im} denote the number of cases in the i^{th} row category, $i = 1, 2, \cdots I$, where the column category is unknown, and let x_{mj} denote the number of cases in the j^{th} column category, $j = 1, 2, \cdots J$ where the row category is unknown. Then, x_{im} and x_{mj} are marginally observed counts on a single variable. Let x_{mm} denote the number of cases where both the row and column categories are missing. The total sample size is

$$
\begin{aligned}
N &= \sum_{ij} x_{ij} + \sum_{i} x_{im} + \sum_{j} x_{mj} + x_{mm} \\
&= x_{cc} + x_{\bullet m} + x_{m\bullet} + x_{mm}.
\end{aligned}
$$

Discarding the x_{mm} cases for which both variables are missing does not affect any results in this paper except that it necessitates changing N to $n = N - x_{mm}$. Those cases do not contain any information about the joint distribution or marginal distributions of X_1 and X_2.

The counts in a completed table obtained by fully efficient fractional imputation under simple random sampling proposed by Kang et al. (2006) are given by the following formula:

$$
\begin{aligned}
\hat{n}^*_{ij} &= x_{ij} + x_{ij}\left(\frac{x_{im}}{x_{i\bullet}} + \frac{x_{mj}}{x_{\bullet j}}\right) \\
&\quad + \frac{x_{ij}x_{mm}}{N - x_{mm}}\left(1 + \frac{x_{im}}{x_{i\bullet}} + \frac{x_{mj}}{x_{\bullet j}}\right) \\
&= x_{ij}\left(1 + \frac{x_{im}}{x_{i\bullet}} + \frac{x_{mj}}{x_{\bullet j}}\right)\left(1 + \frac{x_{mm}}{N - x_{mm}}\right), \quad (1)
\end{aligned}
$$

where x_{ij} is the observed count for fully observed cases prior to imputation, $x_{i\bullet} = \sum_{j=1}^{J} x_{ij}$ and $x_{\bullet j} = \sum_{i=1}^{I} x_{ij}$. The total sample size is $N = \sum_{ij} \hat{n}^*_{ij}$.

When the count x_{mm} is discarded, (1) can be simplified as

$$
\hat{n}_{ij} = x_{ij}\left(1 + \frac{x_{im}}{x_{i\bullet}} + \frac{x_{mj}}{x_{\bullet j}}\right). \quad (2)
$$

The population proportions are estimated as

$$
\begin{aligned}
\hat{\theta}_F &= \frac{1}{n}(\hat{n}_{11}, \hat{n}_{12}, \cdots, \hat{n}_{1J}, \hat{n}_{21}, \hat{n}_{22}, \cdots, \\
&\qquad \hat{n}_{2J}, \cdots, \hat{n}_{IJ})' \quad (3) \\
&= \frac{1}{N}(\hat{n}^*_{11}, \hat{n}^*_{12}, \cdots, \hat{n}^*_{1J}, \hat{n}^*_{21}, \hat{n}^*_{22}, \cdots, \\
&\qquad \hat{n}^*_{2J}, \cdots, \hat{n}^*_{IJ})'.
\end{aligned}
$$

The log-likelihood function for the cell probabilities θ presented by Chen and Fienberg (1974) is the following;

$$
l(\theta) = \sum_{i}\sum_{j} x_{ij}\log\theta_{ij} + \sum_{j} x_{mj}\log\theta_{\bullet j} + \sum_{i} x_{im}\log\theta_{i\bullet}. \quad (4)
$$

Note that the log-likelihood function in (4) does not include x_{mm}.

3. Maximum Likelihood Estimation Of Cell Probabilities And Variances

The EM algorithm (Dempster, Laird, and Rubin 1977) can be used to produce maximum likelihood estimates (MLEs) of proportions in an incomplete contingency table. Little (1982) includes an illustration of the procedure with partially classified data on two variables. Let $\theta^{(0)}_{ij}$ be an initial estimate of θ_{ij}, such as x_{ij}/x_{cc}. The estimate of θ_{ij} at the t^{th} iteration of the algorithm is

$$
\theta^{(t)}_{ij} = 1/n\left(x_{ij} + x_{im} \times \frac{\theta^{(t-1)}_{ij}}{\theta^{(t-1)}_{i\bullet}} + x_{mj} \times \frac{\theta^{(t-1)}_{ij}}{\theta^{(t-1)}_{\bullet j}}\right). \quad (5)
$$

The estimates at the first step of the EM procedure, $\theta^{(1)}$, is exactly same as $\hat{\theta}_F$ in (4). The algorithm converges to the MLEs of θ.

For the example by Little (1982), the 30 partially classified units with $X_1 = 1$ have $X_2 = 1$ with probability $100/(100 + 50)$ and $X_2 = 2$ with probability $50/(100 + 50)$. Thus in effect, $(30)(100)/150 = 20$ are allocated to $X_2 = 1$ and $(30)(50)/150 = 10$ are allocated to $X_2 = 2$ at the first step. In the next step new cell probabilities are computed from the completed data and the procedure iterates to convergence.

Since the proportions are constrained to sum to one ($\sum_{ij} \theta_{ij} = 1$), the likelihood function (4) incorporating the constraint can be expressed with a Lagrangian multiplier as

$$l(\theta^*) = \sum_i \sum_j x_{ij} \log \theta_{ij} + \sum_j x_{mj} \log \theta_{\bullet j} \quad (6)$$

$$+ \sum_i x_{im} \log \theta_{i\bullet} + \gamma(1 - \sum_{ij} \theta_{ij}), \quad (7)$$

where $\theta^* = (\theta', \gamma)'$.

The first derivative of $l(\theta^*)$ in 6 with respective to θ_{ij} and to γ are

$$\frac{\partial l(\theta^*)}{\partial \theta_{ij}} = \frac{x_{ij}}{\theta_{ij}} + \frac{x_{mj}}{\theta_{\bullet j}} + \frac{x_{im}}{\theta_{i\bullet}} - \gamma$$

$$\frac{\partial l(\theta^*)}{\partial \gamma} = 1 - \sum_{ij} \theta_{ij}.$$

The second partial derivatives are

$$\frac{\partial^2 l(\theta^*)}{\partial \theta_{ij}^2} = \frac{-x_{ij}}{\theta_{ij}^2} - \frac{x_{mj}}{\theta_{\bullet j}^2} - \frac{x_{im}}{\theta_{i\bullet}^2},$$

$$\frac{\partial^2 l(\theta^*)}{\partial \theta_{ij}\partial \theta_{is}} = -\frac{x_{im}}{\theta_{i\bullet}^2}, \quad \frac{\partial^2 l(\theta^*)}{\partial \theta_{ij}\partial \theta_{kj}} = -\frac{x_{mj}}{\theta_{\bullet j}^2}$$

$$\frac{\partial^2 l(\theta^*)}{\partial \theta_{ij}\partial \theta_{ks}} = 0, \quad \frac{\partial^2 l(\theta^*)}{\partial \theta_{ij}\partial \gamma} = -1, \quad \frac{\partial^2 l(\theta^*)}{\partial \gamma^2} = 0$$

Let $I(\theta^*)$ be the matrix of second derivatives of the loglikelihood with respect to θ^*. An estimate of the covariance matrix of $\hat{\theta}^*$, the MLE of θ^*, is $-I^{-1}(\hat{\theta}^*)$. Let $\hat{\Sigma}_M$ be a matrix omitting the last row and column of $-I^{-1}(\hat{\theta}^*)$. The matrix $\hat{\Sigma}_M$ gives an estimate of the covariance matrix of $\hat{\theta}_M$, where $\hat{\theta}_M$ is the MLE of θ. For a 2×2 table, $\theta^* = (\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}, \gamma)'$ and $I(\theta^*)$ is

$$I(\theta^*) = -\begin{pmatrix} d_{11} & \frac{x_{1m}}{\theta_{1\bullet}^2} & \frac{x_{m1}}{\theta_{\bullet 1}^2} & 0 & 1 \\ \frac{x_{1m}}{\theta_{1\bullet}^2} & d_{12} & 0 & \frac{x_{m2}}{\theta_{\bullet 2}^2} & 1 \\ \frac{x_{m1}}{\theta_{\bullet 1}^2} & 0 & d_{21} & \frac{x_{2m}}{\theta_{2\bullet}^2} & 1 \\ 0 & \frac{x_{m2}}{\theta_{\bullet 2}^2} & \frac{x_{2m}}{\theta_{2\bullet}^2} & d_{22} & 1 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix},$$

where $d_{ij} = \frac{x_{ij}}{\theta_{ij}^2} + \frac{x_{mj}}{\theta_{\bullet j}^2} + \frac{x_{im}}{\theta_{i\bullet}^2}$.

4. Tests of Independence

A Wald test and a likelihood ratio test for independence in an incomplete two-way contingency table are described in this section.

4.1 A Wald Test Using MLE in an Incomplete Contingency Table

For a complete two-dimensional contingency table with sample size N, the null hypothesis of statistical independence is

$$H_0 : \theta_{ij} = \theta_{i\bullet}\theta_{\bullet j}, \ \text{ for all } i \text{ and } j. \quad (8)$$

Defining

$$g_{ab}(\theta) \equiv \left(\sum_{j=1}^J \theta_{aj}\right)\left(\sum_{i=1}^I \theta_{ib}\right) - \theta_{ab},$$

the null hypothesis of statistical independence can be expressed as

$$H_0 : g_{ab}(\theta) = 0 \quad (9)$$

for all $a = 1, 2, \cdots, I$ and $b = 1, 2, \cdots, J$. Let

$$g(\theta) = (g_{11}(\theta), \cdots, g_{1J}(\theta), g_{21}(\theta), \cdots, g_{2J}(\theta), \cdots, g_{IJ}(\theta))'.$$

Then (9) can be expressed as $H_0 : g(\theta) = 0$.

The estimator, $\hat{\theta}$, has an approximate multivariate normal distribution with variance $\Sigma = V_\theta/N$ by the Central Limit Theorem, where $V_\theta = (\Delta_\theta - \theta\theta')$ and Δ_θ is a diagonal matrix with the elements of θ on the main diagonal. Under H_0, $g(\hat{\theta})$ has an approximate $p = I \times J$ dimensional normal distribution with variance $G\Sigma G'$, where $G_{p\times p}$ is the matrix of first partial derivatives of $g(\theta)$ with respect to θ_{ij}. The elements of $G_{p\times p}$ are

$$\frac{\partial g_{ab}(\theta)}{\partial \theta_{ij}} = \begin{cases} \sum_{i=1}^I \theta_{ib} + \sum_{j=1}^J \theta_{aj} - 1, & \text{for } a = i, \text{ and } b = j \\ \sum_{i=1}^I \theta_{ib}, & \text{for } a = i, \text{ and } b \neq j \\ \sum_{j=1}^J \theta_{aj}, & \text{for } a \neq i, \text{ and } b = j \\ 0, & \text{for } a \neq i, \text{ and } b \neq j \end{cases}.$$

A Wald statistic for testing H_0 is

$$\hat{Q} = g(\hat{\theta})'\hat{T}^- g(\hat{\theta}), \quad (10)$$

where $\hat{T} = (\hat{G}\hat{\Sigma}\hat{G}')$ is obtained by substituting $\hat{\theta}$ for θ and $^-$ denotes generalized matrix inverse. For a complete table, $\hat{Q}$ has a distribution converging to a central chi-squared distribution with $df = k = (I - 1)(J - 1)$ when H_0 is true.

For an incomplete contingency table, $g(\hat{\theta})$ and $\hat{G}$ are obtained by substituting $\hat{\theta}_M$ for $\hat{\theta}$. Then $\hat{T} = (\hat{G}\hat{\Sigma}\hat{G}')$ is obtained by substituting $\hat{\Sigma}_M$ for $\hat{\Sigma}$ in (10). Thus a Wald statistic for testing H_0 is

$$\hat{Q}_M = g(\hat{\theta}_M)'\hat{T}_M^- g(\hat{\theta}_M), \quad (11)$$

where $\hat{T}_M = (\hat{G}\hat{\Sigma}_M\hat{G}')$ and $\hat{G}$ is obtained by substituting $\hat{\theta}_M$ for θ. Then for a incomplete table, $\hat{Q}_M$ has an approximate central chi-square distribution with $df = k = (I-1)(J-1)$ when H_0 is true.

4.2 Likelihood Ratio Test

Under the model of independence in (8), the MLE of θ_{ij} is

$$\hat{\theta}^0_{ij} = \left(\frac{x_{i\bullet} + x_{im}}{x_{cc} + x_{\bullet m}}\right)\left(\frac{x_{\bullet j} + x_{mj}}{x_{cc} + x_{m\bullet}}\right). \quad (12)$$

Let L_0 be the maximized value of the log-likelihood function under the null hypothesis of (8) and L_1 be the maximized value under $H_0 \bigcup H_a$. L_0 can be calculated directly by substituting $\hat{\theta}^0_{ij}$ for θ_{ij} in (6) and L_1 can be obtained by using $\hat{\theta}_M$. Then $2L_1 - 2L_0$ has a limiting null chi-squared distribution as n goes to infinity when H_0 is true. A size α test is implemented by rejecting H_0 if $2L_1 - 2L_0 > \chi^2_{k,\alpha}$.

5. Simulation to Compare Three Methods

The performance of tests of independence on incomplete tables using maximum likelihood estimation (MLE), multiple imputation (MI), and complete case analysis (CC) are compared through Monte Carlo simulations. Two missing data mechanisms corresponding to the missing at random (MAR) assumption are used in simulations. Type I error levels are estimated from 1,000 tables simulated tables under the independence assumption. Power levels are examined by simulating 1000 tables under an alternative to independence.

For multiple imputation, the Wald statistic proposed by Li, Raghunathan, and Rubin (1991) and the likelihood ratio test statistic proposed by Meng and Rubin(1992) based on five imputed data sets were applied to test independence. The algorithms for MI were programmed through S-PLUS 6.1 (2001) functions for missing values.

5.1 Type I Error Levels

The 2×2 incomplete contingency tables generated for this study to check type I error level were generated with equal cell probabilities and data missing at random(MAR). X_1 and X_2 were independently generated as Bernouli(0.5) random variables with sample size 500.

There are two cases with different missing at random mechanisms; 1,000 tables were generated for each case. The missing mechanism for each case is as follows:

$$\begin{aligned} Pr(X_1 \text{ is missing}|X_2 = 1) &= m_1 \quad (13) \\ Pr(X_1 \text{ is missing}|X_2 = 0) &= m_2 \\ Pr(X_2 \text{ is missing}) &= m_3 \end{aligned}$$

For the first case $m_1 = 0.1$, $m_2 = 0.3$, $m_3 = 0.2$ in (13). In the second case $m_1 = 0.2$, $m_2 = 0.4$, $m_3 = 0.3$. The percentages of cases with missing information on at least one variable are expected to be 36% and 51% for case 1 and 2 respectively.

Table 1 shows the numbers of tables for which the independence null hypothesis was falsely rejected out of 1000 tables for three nominal Type I error levels. The results using MLE and CC seem to have appropriate Type I error levels on both tests, but Type I error levels tend to be inflated for the MI method.

5.2 Power Study

An alternative to independence for 2×2 tables with equal probability margins was used to compare the power of the various procedures. The generated multinomial variables have the cell probabilities

$$(\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}) = (0.2, 0.3, 0.3, 0.2), \quad (14)$$

with sample size 500. The missing data mechanisms (13) are same as for the two cases used previously.

Before we study power of independence test, let's compare three methods with checking point estimations of θ_{11} and $\theta_{1+}\theta_{+1} - \theta_{11}$. Table 2 shows means and standard deviations of 1000 values for the estimates of θ_{11} from the generated 1000 tables in this subsection. The true value of θ_{11} is 0.2. MLE and MI methods provide essentially unbiased estimates for the cell probabilities but CC does not. The standard deviations of the estimates differ across methods. Complete-case analysis provides the estimate of θ_{11} with the largest variance. For all methods, variation increases as the proportion of missing values increases. MLE tends to provide smaller standard deviations of cell proportion than MI.

Table 3 shows means and standard deviations of 1000 simulated values for the estimates of $\theta_{1+}\theta_{+1} - \theta_{11}$, a measure of association between the two variables. The true value of $\theta_{1+}\theta_{+1} - \theta_{11}$ is 0.05. The averages of the estimates are similar for all methods. The complete-case and MLE have similar standard deviations and they exhibit smaller standard deviations than MI. Results on point estimation are helpful in interpreting power simulation results.

The numbers in Table 4 indicate the number of tables out of 1000 for which the independence null hypothesis was rejected under the given α levels among 1000 tables in each type.

Table 4 shows MLE and CC have more power than MI. MLE does not show much improvement on the power levels of the tests of independence over complete-case analysis. Although MLE is often more conservative than MI with respect to Type I error levels, the test using MLE exhibited more power than MI.

6. Summary and Discussion

It has been an issue to estimate the variance of MLE for the cell probabilities of an incomplete contingency table because it is very complicated to get the second derivatives of the likelihood. The likelihood including a Lagrangian multiplier related to a constraint can solve this problem.

Complete-case (CC) analysis produces biased estimates of joint probabilities under MAR and is less efficient than either MLE or MI. MLE and MI provide consistent results under either MAR situation used in simulations.

When data are missing at random, simulation results reveal that MLE provides more efficient estimates of population proportions than either multiple imputation (MI) based on data augmentation or complete case analysis, but neither MLE nor MI provides an improvement over complete-case (CC) analysis with respect to power of tests for independence.

If the missing mechanism does satisfy missing completely at random (MCAR) criterion, CC analysis can produce unbiased estimates of joint probabilities and moderate type I error level and power.

Fully efficient fraction imputation (FEFI) for contingency tables was examined in Kang, Koehler, and Larsen (2006a, 2006b). When data are MCAR, multiple imputation and fully efficient fractional imputation have similar levels of variation in point estimates and confidence interval coverage. There is no evidence that imputation is advantageous for hypothesis testing. Complete case analysis appears to be just as good when the data are missing completely at random, although estimation of parameters is improved by FEFI.

Future work will generalize results to multiway tables and tests related to hierarchically nested log linear models. Cases in which data are not missing at random (NMAR) and nonignorably missing (Little and Rubin 2002) will be studied.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 0532413. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Aitchison, J., and Silvey, S. D. (1958). "Maximum-Likelihood Estimation of Parameters Subject to Restraints." *The Annals of Mathmatical Statistics*, 29, 813-828.

Chen, T. T., and Fienberg, S. E. (1974). "Two -dimensional contingency tables with both completely and partially cross-classified data." *Biometrics*, 32, 133-144.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). "Maximum likelihood from incomplete data via the EM algorithm (with discussion)." *Journal of the Royal Statistical Society, Series B*, 39, 1-38.

Hocking, R. R., and Oxspring, H. H. (1974). "The Analysis of Partially Categorized contingency Data." *Biometrics* , 30, 469-483.

Kang, S. S., Koehler, K. J., and Larsen, M. D. (2006a). "Fractional Imputation and Tests of Independence for Incomplete Two-Way Contingency Tables" *Iowa State University, Department of Statistics Preprint*. Submitted to *Computational Statistics and Data Analysis*.

Kang, S. S., Koehler, K. J., and Larsen, M. D. (2006b). "Fractional Imputation for Incomplete Two-Way Contingency Tables" *Iowa State University, Department of Statistics Preprint*. Submitted to *Communications in Statistics - Theory and Methods*.

Li, K. H., Raghunathan, T. E., and Rubin, D. B. (1991). "Large-Sample Significance Levels from Multiple Imputed Data Using Moment-Based Statistics and an F Reference Distribution." *Journal of the American Statistical Association*, 86, 1065-1073.

Lipsitz, S. R., Parzen, M., and Molenberghs, G. (1998). "Obtaining the maximum likelihood estimates in incomplete $R \times C$ contingency tables using a Poisson generalized model." *Journal of Computational and Graphical Statistics*, 7, 356-376.

Little, R. J. A. (1982). "Models for nonresponse in sample surveys." *Journal of the American Statistical Association*, 77, 237-250.

Little, R. J. A., and Rubin, D. B. (2002). *Statistical analysis with missing data.* J. Wiley & Sons, New York.

Meng, X. L., and Rubin, D. B. (1992). "Performing likelihood ratio tests with multiply-imputed data sets." *Biometrika*, 79, 103-111.

Rubin, D. B. (1976). "Inference and missing data (with discussion)." *Biometrika*, 63, 581-592.

Rubin, D. B. (1978). "Multiple Imputation in Sample Surveys - A Phenomenological Bayesian Approach to Nonresponse ." in *Proceedings of the Survey Research Methods Section, American Statistical Association 1978*, 20-34.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. John Wiley & Sons: New York.

Rubin, D. B. (1996). "Multiple imputation after 18+ years." *Journal of the American Statistical Association*, 91, 473-489.

S-Plus 6.1 Manual: Analyzing Data with Missing Values in S-Plus (2001). Insightful Corporation. Seattle, Washington.

West, C. P., and Dawson, J. D. (2002). "Complete imputation of missing repeated categorical data: one-sample applications." *Statistics in Medicine*, 21, 203-217.

Table 1: **Comparison of Type I Error Levels.**

	Wald test								
	MLE			**MI**			**CC**		
Case \ α	1%	5%	10%	1%	5%	10%	1%	5%	10%
1	10	53	91	10	45	96	9	51	95
2	9	58	109	16	68	110	8	56	109
	Likelihood ratio test								
1	9	52	91	10	44	96	9	51	96
2	7	57	109	16	66	108	8	57	109

Table 2: **Estimation of θ_{11}.**

	MLE		MI		CC	
Case	Mean	S.D.	Mean	S.D.	Mean	S.D.
1	0.1984	0.02063	0.1988	0.02118	0.2242	0.02377
2	0.1978	0.02193	0.1977	0.02281	0.2281	0.02697

Table 3: **Estimation of** $\theta_{1+}\theta_{+1} - \theta_{11}$**.**

	MLE		MI		CC	
Case	Mean	S.D.	Mean	S.D.	Mean	S.D.
1	0.0495	0.01376	0.0493	0.01422	0.0488	0.01358
2	0.0498	0.01608	0.0496	0.01691	0.0490	0.01585

Table 4: **Power Comparison.**

	Wald test								
	MLE			**MI**			**CC**		
Case \ α	1%	5%	10%	1%	5%	10%	1%	5%	10%
1	848	955	976	718	896	953	840	956	976
2	702	874	928	525	765	865	698	876	931
	Likelihood ratio test								
1	846	955	976	709	894	953	840	956	976
2	690	870	928	503	752	862	701	877	932