Statistica Sinica Preprint No: SS-2015-0472R3	
Title	Assessment of nonignorable nonresponse log-linear models for an incomplete two-way contingency table
Manuscript ID	SS-2015-0472R3
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202015.0472
Complete List of Authors	Daeyoung Kim and Seongyong Kim
Corresponding Author	Daeyoung Kim
E-mail	daeyoung@math.umass.edu, kdystat@gmail.com

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0472

Statistica Sinica

ASSESSMENT OF NONIGNORALBE LOG-LINEAR MODELS FOR AN INCOMPLETE CONTINGENCY TABLE

Seongyong Kim and Daeyoung Kim

Hoseo University and University of Massachusetts Amherst

Abstract: A challenging problem in the analysis of an incomplete contingency table is that the use of nonignorable nonresponse models requires explicit specification of missing data mechanism. In this paper we propose a data analytic approach to aid in distinguishing between plausible nonignorable log-linear models for an incomplete contingency table. The proposed method involves the computation of a set of response odds and nonresponse odds that are directly connected with the magnitude of the parameters representing types of nonignorable mechanism assumed in the log-linear models. These odds can be easily estimated from observed counts. We illustrate the performance of the proposed method with simulation and data. We also discuss a generalizability of the proposed method in two directions, its applicability for a three-way incomplete contingency table and its applicability for nonignorable nonresponse models other than the log-linear models.

Key words: Contingency table, Log-linear model, Nonignorable nonresponse.

1. Introduction

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0472

Categorical variables subject to missing data can be summarized in the form of an incomplete contingency table. By the incomplete contingency table we mean a data set with a completely classified table and its supplemental margins summarizing missing data on at least one of the variables. For meaningful inference, one needs to take account of the missing data mechanism.

Missing data mechanism is termed *missing completely at random* (MCAR) if missingness is independent of both observed and unobserved response, *missing at random* (MAR) when missingness depends only on observed response, and *not missing at random* (NMAR) if the missingness depends on both observed and unobserved response. In the likelihood framework MCAR and MAR are *ignorable* in that the statistical inference is valid without explicitly modeling the missing data mechanism. NMAR is termed *nonignorable* as it requires the explicit form of the missing data mechanism (Little and Rubin (2002)).

A challenging issue in using nonignorable nonresponse models is that the types of nonignorable mechanism assumed in the models are not empirically verifiable. Molenberghs, Goetghebeur, Lipsitz, and Kenward (1999) and Molenberghs, Beunckens, Sotto, and Kenward (2008) showed that different nonignorable nonresponse models may provide the same fit to the

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0472

observed data, but give different prediction of unobserved data. To aid the model selection and the assessment of untestable assumptions for the missing data mechanism, several methods for sensitivity analysis have been proposed (Copas and Eguchi (2001); Molenberghs, Kenward, and Goetghebeur (2001); Baker, Ko, and Graubard (2003); Troxel, Ma, and Heitjan (2004); Vansteelandt, Goetghebeur, Kenward, and Molenberghs (2006); Xie, Qian, and Qu (2011)).

To incorporate missing data mechanism for the analysis of incomplete two-way contingency tables, two main approaches have been used: the selection model (Fay (1986); Baker and Laird (1988); Little and Rubin (2002)) and the pattern mixture model (Little (1993); Park and Brown (1994); Foster and Smith (1998)).

As an alternative approach, a family of nonignorable nonresponse models based on a log-linear parameterization, called nonignorable log-linear models, was proposed (Baker and Laird (1988); Baker, Rosenberger, and Dersimonian (1992)). It was shown that the log-linear model for an incomplete two-way table is equivalent to the selection model when only one variable is subject to missing data (Clarke and Smith (2004)), and it has the aspects of both selection model and pattern mixture model when both variables are subject to missing values (Molenberghs, Goetghebeur, Lipsitz,

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0472

and Kenward (1999)).

In this paper we propose a data-analytic method to perform model selection within a family of nonignorable log-linear models given by Baker, Rosenberger, and Dersimonian (1992) for an incomplete two-way contingency table. We develop the measures directly associated with the magnitude of the parameters for the types of nonignorable mechanism assumed in the nonresponse models. These measures are based on a set of response odds computed from the completely classified table and nonresponse odds computed from the supplemental margins. We prove that, for log-linear models describing differential missingness of one variable by the other variable, the inequalities relating nonresponse odds to the range of counterpart response odds indicate the presence of the parameters representing such informative missingness.

We show that the inequalities based on the proposed measures can be estimated using only the observed counts. Thus, without fitting the nonignorable log-linear models to the data, they can be used as a data analytic aid to distinguish between plausible types of nonignorable missingness assumed in the models.

The rest of the paper is organized as follows. Section 2 specifies the eight nonignorable log-linear models for an incomplete two-way contingency

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0472

table. Section 3 presents the theoretical properties of the log-linear models with respect to the proposed measures. Section 4 proposes an easy-to-use method to aid in the assessment of the nonignorable log-linear models. In Section 5 the performance of the proposed method is illustrated with simulations and data example. Section 6 discusses the generalizability of the proposed method in two directions, its applicability for a three-way incomplete table and its applicability for nonignorable nonresponse models other than the nonignorable log-linear models.

2. Nonignorable multinomial log-linear models

Let Y_1 and Y_2 be the row and column variables with I and J categories, respectively. Let R_1 and R_2 be the indicators of missingness for Y_1 and Y_2, respectively, with $R_i = 1$ if Y_i is observed and $R_i = 2$ otherwise, i=1,2. For the full array of Y_1, Y_2, R_1, and R_2, we have an $I \times J \times 2 \times 2$ table with the cell counts $\mathbf{y} = \{y_{ijk\ell}\}$ and the cell probabilities $\boldsymbol{\pi} = \{\pi_{ijk\ell}\}$ where $i = 1, \cdots, I$, $j = 1, \cdots, J$, $k, \ell = 1, 2$ and $\pi_{ijk\ell} = Pr(Y_1 = i, Y_2 = j, R_1 = k, R_2 = \ell)$. What we observe is the incomplete $I \times J$ table shown in Table 1, with $\mathbf{y}_{obs} = (\{y_{ij11}\}, \{y_{i+12}\}, \{y_{+j21}\}, \{y_{++22}\})$, where "+" in the subscripts denotes summation over the corresponding subscript.

Assume that the observed counts $\mathbf{y}_{obs}$ in Table 1 are a realization of a multinomial distribution with the cell probabilities $\boldsymbol{\pi}$ and a fixed total

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0472

Table 1: Incomplete $I \times J$ table

		$R_2 = 1$			$R_2 = 2$
		$Y_2 = 1$	$\cdots$	$Y_2 = J$	Y_2
	$Y_1 = 1$	y_{1111}	$\cdots$	y_{1J11}	y_{1+12}
$R_1 = 1$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
	$Y_1 = I$	y_{I111}	$\cdots$	y_{IJ11}	y_{I+12}
$R_1 = 2$	Y_1	y_{+121}	$\cdots$	y_{+J21}	y_{++22}

count $N=\sum_{i,j,k,\ell} y_{ijk\ell}$. Then the log-likelihood is

$$\ell = \sum_{i=1}^{I}\sum_{j=1}^{J} y_{ij11}\log\pi_{ij11} + \sum_{i=1}^{I} y_{i+12}\log\pi_{i+12} + \sum_{j=1}^{J} y_{+j21}\log\pi_{+j21} + y_{++22}\log\pi_{++22},$$

where $\pi_{ijk\ell} = m_{ijk\ell}/\sum_{i,j,k,\ell} m_{ijk\ell}$ and $\mathbf{m} = \{m_{ijk\ell}\}$ is the matrix of the expected cell counts.

To analyze Table 1 under the nonignorable mechanism, Baker, Rosenberger, and Dersimonian (1992) considered the multinomial log-linear models, as shown in Table 2. The subscript of each λ-term in the models indicates the variable(s) of the (main effect/interaction) parameter and the superscript the level of a variable shown in the subscript. The sum of each λ-term over each of the superscripts is constrained to be zero for identification purposes.

The models in Table 2 differ by the types of nonignorable mechanism characterized by interactions between (Y_1,Y_2) and (R_1,R_2). To better understand them, Baker, Rosenberger, and Dersimonian (1992) used the pa-

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0472

Table 2: Nonignorable multinomial log-linear models for Table 1

Notation	Model
$(\alpha_{i\cdot},\beta_{i\cdot})$	$\log m_{ijk\ell}=\lambda^i_{Y_1}+\lambda^j_{Y_2}+\lambda^k_{R_1}+\lambda^\ell_{R_2}+\lambda^{ij}_{Y_1Y_2}+\lambda^{k\ell}_{R_1R_2}+\lambda^{ik}_{Y_1R_1}+\lambda^{i\ell}_{Y_1R_2}$
$(\alpha_{\cdot j},\beta_{\cdot j})$	$\log m_{ijk\ell}=\lambda^i_{Y_1}+\lambda^j_{Y_2}+\lambda^k_{R_1}+\lambda^\ell_{R_2}+\lambda^{ij}_{Y_1Y_2}+\lambda^{k\ell}_{R_1R_2}+\lambda^{jk}_{Y_2R_1}+\lambda^{j\ell}_{Y_2R_2}$
$(\alpha_{i\cdot},\beta_{\cdot j})$	$\log m_{ijk\ell}=\lambda^i_{Y_1}+\lambda^j_{Y_2}+\lambda^k_{R_1}+\lambda^\ell_{R_2}+\lambda^{ij}_{Y_1Y_2}+\lambda^{k\ell}_{R_1R_2}+\lambda^{ik}_{Y_1R_1}+\lambda^{j\ell}_{Y_2R_2}$
$(\alpha_{\cdot j},\beta_{i\cdot})$	$\log m_{ijk\ell}=\lambda^i_{Y_1}+\lambda^j_{Y_2}+\lambda^k_{R_1}+\lambda^\ell_{R_2}+\lambda^{ij}_{Y_1Y_2}+\lambda^{k\ell}_{R_1R_2}+\lambda^{jk}_{Y_2R_1}+\lambda^{i\ell}_{Y_1R_2}$
$(\alpha_{\cdot\cdot},\beta_{i\cdot})$	$\log m_{ijk\ell}=\lambda^i_{Y_1}+\lambda^j_{Y_2}+\lambda^k_{R_1}+\lambda^\ell_{R_2}+\lambda^{ij}_{Y_1Y_2}+\lambda^{k\ell}_{R_1R_2}+\lambda^{i\ell}_{Y_1R_2}$
$(\alpha_{\cdot\cdot},\beta_{\cdot j})$	$\log m_{ijk\ell}=\lambda^i_{Y_1}+\lambda^j_{Y_2}+\lambda^k_{R_1}+\lambda^\ell_{R_2}+\lambda^{ij}_{Y_1Y_2}+\lambda^{k\ell}_{R_1R_2}+\lambda^{j\ell}_{Y_2R_2}$
$(\alpha_{i\cdot},\beta_{\cdot\cdot})$	$\log m_{ijk\ell}=\lambda^i_{Y_1}+\lambda^j_{Y_2}+\lambda^k_{R_1}+\lambda^\ell_{R_2}+\lambda^{ij}_{Y_1Y_2}+\lambda^{k\ell}_{R_1R_2}+\lambda^{ik}_{Y_1R_1}$
$(\alpha_{\cdot j},\beta_{\cdot\cdot})$	$\log m_{ijk\ell}=\lambda^i_{Y_1}+\lambda^j_{Y_2}+\lambda^k_{R_1}+\lambda^\ell_{R_2}+\lambda^{ij}_{Y_1Y_2}+\lambda^{k\ell}_{R_1R_2}+\lambda^{jk}_{Y_2R_1}$

rameterization,

$$\alpha_{ij}=\frac{\pi_{ij21}}{\pi_{ij11}},\ \beta_{ij}=\frac{\pi_{ij12}}{\pi_{ij11}}. \tag{2.1}$$

The expressions of α_{ij} and β_{ij} are different for each model in Table 2, as shown in Section S1 of the Supplementary Material: α_{ij} is concerned with the dependence of R_1 on Y_1 and Y_2, while β_{ij} is associated with the dependence of R_2 on Y_1 and Y_2. For example, for the models with $\lambda^{ik}_{Y_1R_1}$, α_{ij} depends on only the subscript i ($\alpha_{ij}=\alpha_{i\cdot}$) and, for the models with $\lambda^{jk}_{Y_2R_1}$, α_{ij} is dependent only on j ($\alpha_{ij}=\alpha_{\cdot j}$). For the models with neither $\lambda^{ik}_{Y_1R_1}$ nor $\lambda^{jk}_{Y_2R_1}$, α_{ij} is independent of both i and j ($\alpha_{ij}=\alpha_{\cdot\cdot}$). Similarly, $\beta_{ij}=\beta_{i\cdot}$ for models with $\lambda^{i\ell}_{Y_1R_2}$, $\beta_{ij}=\beta_{\cdot j}$ for the models with $\lambda^{j\ell}_{Y_2R_2}$, and $\beta_{ij}=\beta_{\cdot\cdot}$ for the models with none of $\lambda^{i\ell}_{Y_1R_2}$ and $\lambda^{j\ell}_{Y_2R_2}$. As shown in the first column of Table 2, each model is denoted by the combination of the forms of α_{ij} and β_{ij}.

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0472

For notational simplicity, we take $(\alpha_{i\cdot},\ \beta_{\square})$ and $(\alpha_{\cdot j},\ \beta_{\square})$ as a group of models with $\lambda^{ik}_{Y_1R_1}$ and $\lambda^{jk}_{Y_2R_1}$, respectively, where $\beta_{\square}$ can be one of $\beta_{\cdot\cdot}$, $\beta_{i\cdot}$, $\beta_{\cdot j}$. Similarly, $(\alpha_{\square},\ \beta_{i\cdot})$ and $(\alpha_{\square},\ \beta_{\cdot j})$ denote a group of models with $\lambda^{i\ell}_{Y_1R_2}$ and $\lambda^{j\ell}_{Y_2R_2}$, respectively, where $\alpha_{\square}$ can be one of $\alpha_{\cdot\cdot}$, $\alpha_{i\cdot}$, $\alpha_{\cdot j}$.

Baker, Rosenberger, and Dersimonian (1992) also discussed the identifiability of the log-linear models in Table 2. When $I = J$, all eight models are identifiable and the first four models are saturated. If $I > J$, the models $(\alpha_{i\cdot},\ \beta_{\square})$ are unidentifiable. For a case of $I < J$, the models $(\alpha_{\square},\ \beta_{\cdot j})$ are unidentifiable.

Remark 1. One can consider the ignorable nonresponse models for Table 1 as the important baseline models (Rubin (1976); Copas and Eguchi (2001); Xie, Qian, and Qu (2011)). The MCAR model is represented as the log-linear model without any interaction between (Y_1, Y_2) and (R_1, R_2) : $\log m_{ijk\ell} = \lambda^{i}_{Y_1} + \lambda^{j}_{Y_2} + \lambda^{k}_{R_1} + \lambda^{\ell}_{R_2} + \lambda^{ij}_{Y_1Y_2} + \lambda^{k\ell}_{R_1R_2}$. For the MCAR log-linear model, α_{ij} and β_{ij} in (2.1) are independent of i and j (we denote the MCAR model as $(\alpha_{\cdot\cdot}, \beta_{\cdot\cdot})$). No MAR model for Table 1 can be represented as a log-linear model with interaction between (Y_1, Y_2) and (R_1, R_2), as such log-linear model cannot satisfy the MAR conditions given in Molenberghs, Goetghebeur, Lipsitz, and Kenward (1999). For details, see Section S2 in the Supplementary Material.

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0472

3. Properties of nonignorable log-linear models

In this section we propose measures to provide a theoretical basis for distinguishing between the different nonignorable log-linear models in Table 1. The proposed measures are based on a set of response odds obtained from the completely classified table and nonresponse odds from the supplemental margins.

Given a pair (j, j') of Y_2, define the response odds $\nu_i(j, j')$ to be a ratio of probabilities of column j and j' within the i-th row of the completely classified table (R_1=R_2=1) and the nonresponse odds $\nu(j, j')$ to be a ratio of probabilities of column j and j' in the supplemental column margin (R_1= 2, R_2=1):

$$\nu_i(j, j') = \frac{\pi_{ij11}}{\pi_{ij'11}}, \quad \nu(j, j') = \frac{\pi_{+j21}}{\pi_{+j'21}}. \tag{3.1}$$

We define the response odds intervals for $\nu_i(j, j')$ in (3.1) by

$$OI^{\nu}(j, j') = \left(\nu_n(j, j'), \nu_m(j, j')\right), \tag{3.2}$$

where $\nu_n(j, j')$=$\min_i \nu_i(j, j')$ and $\nu_m(j, j')$=$\max_i \nu_i(j, j')$ are the minimum and maximum value of $\nu_i(j, j')$ over all i at each pair (j, j') of Y_2, respectively.

Similarly, for a given pair (i, i') of Y_1, we define odds as

$$\omega_j(i, i') = \frac{\pi_{ij11}}{\pi_{i'j11}}, \quad \omega(i, i') = \frac{\pi_{i+12}}{\pi_{i'+12}}, \tag{3.3}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0472

where the response odds $\omega_j(i,i')$ is a ratio of probabilities of rows i and i' within column j of the completely classified table and the nonresponse odds $\omega(i,i')$ is a ratio of probabilities rows i and i' in the supplemental row margin (R_1=1,R_2=2). We denote the response odds intervals for $\omega_j(i,i')$ in (3.3) by

$$OI^{\omega}(i,i') = (\omega_n(i,i'), \omega_m(i,i'))\,, \tag{3.4}$$

where $\omega_n(i,i')$=$\min_j \omega_j(i,i')$ and $\omega_m(i,i')$=$\max_j \omega_j(i,i')$ are the minimum and maximum value of $\omega_j(i,i')$ over all j at each pair (i,i') of Y_1, respectively.

The inequalities relating nonresponse odds ($\nu(j,j')$ in (3.1) and $\omega(i,i')$ in (3.3)) to the counterpart response odds intervals ($OI^{\nu}(j,j')$ in (3.2) and $OI^{\omega}(i,i')$ in (3.4)) are associated with the magnitude of the interaction parameters representing types of nonignorable mechanism assumed in the log-linear models.

Theorem 1. *Suppose that $\boldsymbol{\pi} = \{\pi_{ijk\ell}\}$ for an $I \times I \times 2 \times 2$ table is modeled by*

1) the models with $\lambda^{jk}_{Y_2R_1}$, $(\alpha_{\cdot j},\ \beta_{\Box})$.

Then, one and only one of the following must hold for each pair (j,j')

of Y_2 *:*

(1) $\nu(j,j') \in OI^{\nu}(j,j')$ *if and only if*

$$-0.5 \log M_m^{\nu}(j,j') < \lambda_{Y_2R_1}^{j'2} - \lambda_{Y_2R_1}^{j2} < -0.5 \log M_n^{\nu}(j,j'),$$

(2) $\nu(j,j') \notin OI^{\nu}(j,j')$ *if and only if*

$$\lambda_{Y_2R_1}^{j'2} - \lambda_{Y_2R_1}^{j2} < -0.5 \log M_m^{\nu}(j,j') \quad or \quad \lambda_{Y_2R_1}^{j'2} - \lambda_{Y_2R_1}^{j2} > -0.5 \log M_n^{\nu}(j,j'),$$

where $M_m^{\nu}(j,j') = \nu_m(j,j')/\nu(j,j') > 1$, $M_n^{\nu}(j,j') = \nu_n(j,j')/\nu(j,j') < 1$ *in the absence of all* $\lambda_{Y_2R_1}^{jk}$*'s, and they are independent of all* $\lambda_{Y_2}^{j}$*'s and* $\lambda_{R_1}^{k}$*'s.*

2) the models without $\lambda_{Y_2R_1}^{jk}$, $(\alpha_{i\cdot},\ \beta_{\square})$, $(\alpha_{\cdot\cdot},\ \beta_{i\cdot})$, $(\alpha_{\cdot\cdot},\ \beta_{\cdot j})$.

Then, $\nu(j,j') \in OI^{\nu}(j,j')$ *for any given pair* (j,j') *of* Y_2.

3) the models with $\lambda_{Y_1R_2}^{i\ell}$, $(\alpha_{\square},\ \beta_{i\cdot})$.

Then, one and only one of the following must hold for each pair (i,i') *of* Y_1 *:*

(1) $\omega(i,i') \in OI^{\omega}(i,i')$ *if and only if*

$$-0.5 \log M_m^{\omega}(i,i') < \lambda_{Y_1R_2}^{i'2} - \lambda_{Y_1R_2}^{i2} < -0.5 \log M_n^{\omega}(i,i'),$$

(2) $\omega(i,i') \notin OI^{\omega}(i,i')$ *if and only if*

$$\lambda_{Y_1R_2}^{i'2} - \lambda_{Y_1R_2}^{i2} < -0.5 \log M_m^{\omega}(i,i') \quad or \quad \lambda_{Y_1R_2}^{i'2} - \lambda_{Y_1R_2}^{i2} > -0.5 \log M_n^{\omega}(i,i'),$$

where $M_m^{\omega}(i,i') = \omega_m(i,i')/\omega(i,i') > 1,\ M_n^{\omega}(i,i') = \omega_n(i,i')/\omega(i,i') <$

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0472

1 *in the absence of all* $\lambda^{i\ell}_{Y_1R_2}$*'s, and they are independent of all* $\lambda^{i}_{Y_1}$*'s and* $\lambda^{\ell}_{R_2}$*'s.*

4) the models without $\lambda^{i\ell}_{Y_1R_2}$, $(\alpha_{\square},\ \beta_{\cdot j})$, $(\alpha_{i\cdot},\ \beta_{\cdot\cdot})$, $(\alpha_{\cdot j},\ \beta_{\cdot\cdot})$.

Then, $\omega(i,i') \in OI^{\omega}(i,i')$ *for any given pair* (i,i') *of* Y_1.

Proof. See Appendix A.1. □

Theorem 1-1) shows that, for the models with $\lambda^{jk}_{Y_2R_1}$ $(\alpha_{\cdot j},\ \beta_{\square})$, the existence of at least one pair (j,j') of Y_2 satisfying $\nu(j,j') \notin OI^{\nu}(j,j')$ is equivalent to the existence of a subset of parameter space such that the value of $\mid \lambda^{j'2}_{Y_2R_1} - \lambda^{j2}_{Y_2R_1} \mid$ is far from zero and thus the presence of a strong interaction between Y_2 and R_1. On the other hand, Theorem 1-2) shows that, for the models without $\lambda^{jk}_{Y_2R_1}$, $\nu(j,j') \in OI^{\nu}(j,j')$ at all pairs (j,j') of Y_2.

In the same way, by Theorem 1-3) and 4), for the models with $\lambda^{i\ell}_{Y_1R_2}$ $(\alpha_{\square},\ \beta_{i\cdot})$, if there exist at least one pair (i,i') of Y_1 satisfying $\omega(i,i') \notin OI^{\omega}(i,i')$, then the models $(\alpha_{\square},\ \beta_{i\cdot})$ have a strong interaction between Y_1 and R_2, compared to the log-linear models with zero values of $\lambda^{i\ell}_{Y_1R_2}$'s, $(\alpha_{\square},\ \beta_{\cdot j})$, $(\alpha_{i\cdot},\ \beta_{\cdot\cdot})$, and $(\alpha_{\cdot j},\ \beta_{\cdot\cdot})$.

These observations have an important implication: the inequalities relating nonresponse odds to the counterpart response odds intervals are useful in distinguishing the models with the interactions describing differential missingness of one variable by the other variable, $(\alpha_{\cdot j}, \beta_{\square})$ with $\lambda^{jk}_{Y_2R_1}$ and $(\alpha_{\square}, \beta_{i\cdot})$ with $\lambda^{i\ell}_{Y_1R_2}$.

One can view the proposed inequalities as a tool to identify missingness in an incomplete $I \times J$ table when there is substantial discrepancy between a supplemental (column/row) margin and the corresponding margin of the completely classified table; such substantial discrepancy indicates that the categories of the dimension over which these margins collapse are weighted dissimilarly in these two margins. Theorem 1 above formalizes the intuition for such missingness using the proposed inequalities.

Remark 2. For an $I \times J \times 2 \times 2$ table with $I \neq J$, Theorem 1 still holds for the identifiable log-linear models in Table 2. For the MCAR log-linear model $(\alpha_{\cdot\cdot}, \beta_{\cdot\cdot})$ given in Remark 1, it can be shown that $\omega(i, i') \in OI^{\omega}(i, i')$ for any given pair (i, i') of Y_1 and $\nu(j, j') \in OI^{\nu}(j, j')$ for any given pair (j, j') of Y_2. For more details, see Section S2.1 in the Supplementary Material.

4. Assessment of the nonignorable log-linear models

This section proposes a data analytic method based on Theorem 1 to assess the suitability of the models in Table 2 for incomplete $I \times I$ tables.

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0472

Table 3: ML estimators for $\nu_i(j,j')$, $\nu(j,j')$, $\omega_j(i,i')$ and $\omega(i,i')$

Model	$\hat\nu_i(j,j')$	$\hat\nu(j,j')$
$(\alpha_{i\cdot},\ \beta_\square)$, $(\alpha_{\cdot j},\ \beta_\square)$	$\frac{y_{ij11}}{y_{ij'11}}$	$\frac{y_{+j21}}{y_{+j'21}}$
$(\alpha_{\cdot\cdot},\ \beta_{i\cdot})$, $(\alpha_{\cdot\cdot},\ \beta_{\cdot j})$	$\frac{y_{ij11}}{y_{ij'11}}\frac{y_{+j+1}/y_{+j11}}{y_{+j'+1}/y_{+j'11}}$	$\frac{\sum_i y_{ij11}}{\sum_i y_{ij'11}}\frac{y_{+j+1}/y_{+j11}}{y_{+j'+1}/y_{+j'11}}$

Model	$\hat\omega_j(i,i')$	$\hat\omega(i,i')$
$(\alpha_\square,\ \beta_{i\cdot})$, $(\alpha_\square,\ \beta_{\cdot j})$	$\frac{y_{ij11}}{y_{i'j11}}$	$\frac{y_{i+12}}{y_{i'+12}}$
$(\alpha_{i\cdot},\ \beta_{\cdot\cdot})$, $(\alpha_{\cdot j},\ \beta_{\cdot\cdot})$	$\frac{y_{ij11}}{y_{i'j11}}\frac{y_{i+1+}/y_{i+11}}{y_{i'+1+}/y_{i'+11}}$	$\frac{\sum_j y_{ij11}}{\sum_j y_{i'j11}}\frac{y_{i+1+}/y_{i+11}}{y_{i'+1+}/y_{i'+11}}$

For the application of Theorem 1, we first give in Table 3 below the maximum likelihood (ML) estimators of the response odds and the nonresponse odds in (3.1) and (3.3) under the log-linear models in Table 2:

$$\hat\nu_i(j,j') = \frac{\hat\pi_{ij11}}{\hat\pi_{ij'11}}, \quad \hat\nu(j,j') = \frac{\hat\pi_{+j21}}{\hat\pi_{+j'21}}, \tag{4.1}$$

$$\hat\omega_j(i,i') = \frac{\hat\pi_{ij11}}{\hat\pi_{i'j11}}, \quad \hat\omega(i,i') = \frac{\hat\pi_{i+12}}{\hat\pi_{i'+12}}, \tag{4.2}$$

where $\hat\pi_{ij11}$, $\hat\pi_{i+12}$, and $\hat\pi_{+j21}$ are the ML estimators of π_{ij11}, π_{i+12}, and π_{+j21}; their closed forms are provided in Section S3 of the Supplementary Material.

We let the estimators for the response odds intervals $OI^\nu(j,j')$ and $OI^\omega(i,i')$ in (3.2) and (3.4) be denoted by, respectively,

$$\widehat{OI}^\nu(j,j') = (\hat\nu_n(j,j'), \hat\nu_m(j,j'))\,, \quad \widehat{OI}^\omega(i,i') = (\hat\omega_n(i,i'), \hat\omega_m(i,i'))\,, \tag{4.3}$$

where $\hat\nu_n(j,j')$=$\min_i \hat\nu_i(j,j')$, $\hat\nu_m(j,j')$=$\max_i \hat\nu_i(j,j')$, $\hat\omega_n(i,i')$=$\min_j \hat\omega_j(i,i')$

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0472

and $\hat{\omega}_m(i,i')$=$\max_j \hat{\omega}_j(i,i')$.

We can see from Table 3 that $\hat{\nu}_i(j,j')$ and $\hat{\nu}(j,j')$ under the models $(\alpha_{\cdot j},\ \beta_{\square})$ are the same as those under the models $(\alpha_{i\cdot},\ \beta_{\square})$. Thus, by Theorem 1-1) and 2), the existence of at least one pair (j,j') of Y_2 satisfying $\hat{\nu}(j,j') \notin \widehat{OI}^{\nu}(j,j')$ corresponds to the presence of at least one large estimated value of $\mid \lambda^{j2}_{Y_2R_1} \mid$ relative to zero, and thus the plausible models for the data at hand are $(\alpha_{\cdot j},\ \beta_{\square})$, not $(\alpha_{i\cdot},\ \beta_{\square})$. Likewise, $\hat{\omega}_j(i,i')$ and $\hat{\omega}(i,i')$ for the models $(\alpha_{\square},\ \beta_{i\cdot})$ are the same as those for the models $(\alpha_{\square},\ \beta_{\cdot j})$. If there exist at least one pair (i,i') of Y_1 satisfying $\hat{\omega}(i,i') \notin \widehat{OI}^{\omega}(i,i')$ for a given data set, Theorem 1-3) and 4) indicate the existence of at least one large estimated value of $\mid \lambda^{i2}_{Y_1R_2} \mid$ relative to zero, and thus the plausible models are $(\alpha_{\square},\ \beta_{i\cdot})$, not $(\alpha_{\square},\ \beta_{\cdot j})$.

By using the estimators under the models $(\alpha_{\cdot\cdot},\ \beta_{i\cdot})$, $(\alpha_{\cdot\cdot},\ \beta_{\cdot j})$, $(\alpha_{i\cdot},\ \beta_{\cdot\cdot})$ and $(\alpha_{\cdot j},\ \beta_{\cdot\cdot})$ in Table 3, we can easily verify Theorem 1-2) and 4): under $(\alpha_{\cdot\cdot},\ \beta_{i\cdot})$ and $(\alpha_{\cdot\cdot},\ \beta_{\cdot j})$, $\hat{\nu}(j,j') \in \widehat{OI}^{\nu}(j,j')$ for all pairs (j,j') and $\hat{\omega}(i,i') \in \widehat{OI}^{\omega}(i,i')$ for all pairs (i,i') under $(\alpha_{i\cdot},\ \beta_{\cdot\cdot})$ and $(\alpha_{\cdot j},\ \beta_{\cdot\cdot})$.

In general, saturated models often provide the best fit for the observed data and, whenever necessary, also offer useful information in the search for more parsimonious models. Furthermore, as shown in Molenberghs, Beunckens, Sotto, and Kenward (2008), saturated nonignorable log-linear

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0472

models can be used to construct MAR counterparts with the same fit for an incomplete contingency table. We have easy-to-use conditions to aid in assessing the four saturated nonignorable log-linear models in Table 2, without fitting them to observed data.

Corollary 1. *Suppose that the nonresponse odds in* (4.1) *and* (4.2) *and the response odds intervals in* (4.3) *are computed using the estimators in Table 3 under the four saturated models* $(\alpha_{i\cdot},\ \beta_{i\cdot})$, $(\alpha_{\cdot j},\ \beta_{\cdot j})$, $(\alpha_{i\cdot},\ \beta_{\cdot j})$, *and* $(\alpha_{\cdot j},\ \beta_{i\cdot})$. *If*

$$C_1 : \hat{\nu}(j,j') \notin \widehat{OI}^{\nu}(j,j') \text{ for at least one pair } (j,j') \text{ of } Y_2, \text{ and}$$

$$C_2 : \hat{\omega}(i,i') \notin \widehat{OI}^{\omega}(i,i') \text{ for at least one pair } (i,i') \text{ of } Y_1,$$

then, the plausible saturated models are given by

$$\begin{cases} (\alpha_{\cdot j},\ \beta_{i\cdot}) & \text{if both } C_1 \text{ and } C_2 \text{ hold,} \\ (\alpha_{\cdot j},\ \beta_{\cdot j}), (\alpha_{\cdot j},\ \beta_{i\cdot}) & \text{if } C_1 \text{ holds and } C_2 \text{ does not hold,} \\ (\alpha_{i\cdot},\ \beta_{i\cdot}), (\alpha_{\cdot j},\ \beta_{i\cdot}) & \text{if } C_1 \text{ does not hold and } C_2 \text{ holds,} \\ (\alpha_{i\cdot},\ \beta_{i\cdot}), (\alpha_{\cdot j},\ \beta_{\cdot j}), (\alpha_{i\cdot},\ \beta_{\cdot j}), (\alpha_{\cdot j},\ \beta_{i\cdot}) & \text{if neither } C_1 \text{ nor } C_2 \text{ hold.} \end{cases} \tag{4.4}$$

To assess the uncertainty of the accuracy of the proposed conditions in Corollary 1, one can use the bootstrap procedure (Efron and Tibshirani (1993)): generate B bootstrap samples of size N from each of the selected models and compute the percentage of bootstrap samples satisfying each of

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0472

the two conditions under each selected model. When a condition is satisfied (not satisfied) for the original data, the corresponding bootstrap percentage close to 100% (0%) indicates that the result of the condition for the original data is accurate.

In the search of more parsimonious nonignorable log-linear models, one needs to consider nested models where one of two interactions between (Y_1, Y_2) and (R_1, R_2) in the saturated model(s) selected by Corollary 1 is equal to zero, because such nested models also satisfy the properties of Theorem 1 under their respective estimators in Table 3. We summarize the nested log-linear models in Table 2 to be considered according to the results of Corollary 1:

$$\begin{cases} (\alpha_{\cdot\cdot},\ \beta_{i\cdot}), (\alpha_{\cdot j},\ \beta_{\cdot\cdot}) & \text{if both } C_1 \text{ and } C_2 \text{ hold,} \\ (\alpha_{\cdot j},\ \beta_{\cdot\cdot}), (\alpha_{\cdot\cdot},\ \beta_{\cdot j}), (\alpha_{\cdot\cdot},\ \beta_{i\cdot}) & \text{if } C_1 \text{ holds and } C_2 \text{ does not hold,} \\ (\alpha_{\cdot\cdot},\ \beta_{i\cdot}), (\alpha_{i\cdot},\ \beta_{\cdot\cdot}), (\alpha_{\cdot j},\ \beta_{\cdot\cdot}) & \text{if } C_1 \text{ does not hold and } C_2 \text{ holds,} \\ (\alpha_{\cdot\cdot},\ \beta_{\cdot j}), (\alpha_{\cdot\cdot},\ \beta_{i\cdot}), (\alpha_{i\cdot},\ \beta_{\cdot\cdot}), (\alpha_{\cdot j},\ \beta_{\cdot\cdot}) & \text{if neither } C_1 \text{ nor } C_2 \text{ hold.} \end{cases} \tag{4.5}$$

If one is interested in the statistical significance of the interactions between (Y_1, Y_2) and (R_1, R_2) in the selected saturated model(s), one can employ the commonly used model selection criteria such as the likelihood ratio test (LRT), the Akaike information criterion (AIC) (Akaike (1974)) and the Bayesian information criterion (BIC) (Schwarz (1978)) for the selected saturated model(s) and the corresponding nested models.

Remark 3. For an incomplete $I \times J$ table with $I \neq J$, the ML estimators in Table 3 and the proposed conditions in Corollary 1 are available for the identifiable log-linear models, the models except $(\alpha_{i\cdot},\ \beta_{\square})$ when $I > J$ and the models except $(\alpha_{\square}, \beta_{\cdot j})$ when $I < J$. For the MCAR model in Remark 1, the ML estimators of $\pi_{ij11},\ \pi_{i+12},\ \pi_{+j21}$ in (4.1) and (4.2) need to be obtained numerically, as their closed forms are unavailable.

5. Numerical examples

In this section we report on simulations and data analysis to examine the proposed methods in Section 4 ((4.4) in Corollary 1 and (4.5)). In the Supplemental material (Section S5), we present another data example using the proposed method. To verify the validity of the proposed method, we also employ the commonly used model selection criteria, the LRT, AIC, BIC, and the deviance statistic G^2 evaluating the goodness-of-fit of the estimated model. The LRT is performed between a saturated model and its nested models. The G^2 is defined to be $2(\ell_{Full} - \ell_{Fit})$ where ℓ_{Full} is the log likelihood evaluated at $(\pi_{ij11}, \pi_{i+12},\ \pi_{+j21}, \pi_{++22}) = (y_{ij11}, y_{i+12}, y_{+j21}, y_{++22})/N$, and ℓ_{Fit} is the log likelihood evaluated at $(\pi_{ij11}, \pi_{i+12}, \pi_{+j21}, \pi_{++22})$ computed

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0472

from the estimated model. Unlike the proposed method, the use of the LRT, AIC, BIC, and G^2 requires estimation of the models using a numerical algorithm.

5.1 Simulation studies

We performed simulations of the $2 \times 2 \times 2 \times 2$ table where Y_1 and Y_2 are dichotomous and R_1 and R_2 are their missingness indicators. In them, three factors were considered: the sample size (total count) N, the type of nonignorable missingness, and its degree. We considered two levels of sample size, N=(5000,10000). As to the type of nonignorable missingness, we considered the four saturated models in Table 2, $(\alpha_{i\cdot},\ \beta_{i\cdot})$, $(\alpha_{\cdot j},\ \beta_{\cdot j})$, $(\alpha_{i\cdot},\ \beta_{\cdot j})$ and $(\alpha_{\cdot j},\ \beta_{i\cdot})$. Each of them describes different types of missingness represented by the interactions between (Y_1, Y_2) and (R_1, R_2): $(\lambda^{11}_{Y_1R_1}$, $\lambda^{11}_{Y_1R_2})$ for $(\alpha_{i\cdot},\ \beta_{i\cdot})$, $(\lambda^{11}_{Y_2R_1}, \lambda^{11}_{Y_2R_2})$ for $(\alpha_{\cdot j},\ \beta_{\cdot j})$, $(\lambda^{11}_{Y_1R_1}, \lambda^{11}_{Y_2R_2})$ for $(\alpha_{i\cdot},\ \beta_{\cdot j})$, and $(\lambda^{11}_{Y_2R_1}, \lambda^{11}_{Y_1R_2})$ for $(\alpha_{\cdot j},\ \beta_{i\cdot})$. For identifiability, there are only two parameters for interactions between (Y_1, Y_2) and (R_1, R_2) in each saturated model.

The degree of the nonignorable missingness assumed in each model was determined by the magnitudes of two interaction parameters between (Y_1, Y_2) and (R_1, R_2). We used four values for each interaction, (0.05, 0.1, 0.2, 0.4), and so there were 16 pairs of values of two interactions in each

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0472

saturated model. The odds ratios of the four corresponding values of the interaction were 1.22, 1.49, 2.23 and 4.95, respectively. For detailed information on the parameter values used in the simulation study, see Section S4.1 in the Supplementary Material.

The three simulation factors were fully crossed, leading to 128 experimental conditions (= 2 for sample size $\times$ 4 for type of nonignorable missingness $\times$ 16 for a combination of two interactions describing the degree of an assumed missingness). Under each experimental condition, we generated 1000 tables and applied the proposed method in (4.4) of Corollary 1, as well as three commonly used model selection criteria, AIC, BIC, and G^2, to each simulated data set. We then counted the number of cases where each of the four saturated models was chosen as a plausible model by the four model selection criteria including the proposed method. By a plausible model in the use of AIC, BIC, and G^2, we mean a saturated model producing the smallest value of AIC and BIC, giving a perfect fit to the observed cell counts, and providing non-boundary solutions for the estimates of unobserved cell counts (Clarke and Smith (2004)), non-zero values of $\hat{m}_{+j+2} = \sum_{i=1}^{2} \sum_{k=1}^{2} \hat{m}_{ijk2}$ for all j's and $\hat{m}_{i+2+} = \sum_{j=1}^{2} \sum_{\ell=1}^{2} \hat{m}_{ij2\ell}$ for all i's.

Table 4 presents the number of the cases each of the four saturated models was chosen by AIC, BIC, G^2, and the proposed method in (4.4) when 1000 tables of size N=5000 are generated from each of the four types of nonignorable missingness. The simulation results for N=10000 are provided in Section S4.2 of the Supplementary Material. The following observations can be made from the simulation.

The performance of the proposed method in (4.4) was exactly the same as those of AIC, BIC, and G^2 in terms of the selection of saturated models, regardless of the sample size, the type of nonignorable missingness and its degree. Moreover, the true (simulation) model is always chosen to be plausible, regardless of three simulation factors.

Given the sample size, the number of selected saturated models changes depending on types of nonignorable missingness and its degrees. Thus, when the type of nonignorable missingness is characterized by the simulation model $(\alpha_{i\cdot}, \beta_{i\cdot})$, the two saturated models $(\alpha_{i\cdot}, \beta_{i\cdot})$ and $(\alpha_{\cdot j}, \beta_{i\cdot})$ are always chosen. When $\lambda^{11}_{Y_1R_2}$ is small (e.g., 0.05, 0.1), the other two saturated models $(\alpha_{i\cdot}, \beta_{\cdot j})$ and $(\alpha_{\cdot j}, \beta_{\cdot j})$ are additionally chosen in more than half the cases. This result can be explained by Theorem 1-2) and 3). Under the simulation model $(\alpha_{i\cdot}, \beta_{i\cdot})$, $\nu(j, j') \in OI^{\nu}(j, j')$ holds for any given pair (j, j') of Y_2 by Theorem 1-2), so the condition C_1 in Corollary 1 does not hold regardless of the size of $\lambda^{11}_{Y_1R_1}$. On the other hand, as shown in Theorem 1-3), the occurrence of $\omega(i,i') \notin OI^{\omega}(i,i')$ depends on the magnitude of $\lambda^{11}_{Y_1R_2}$, and so does the occurrence of the condition C_2 in Corollary 1.

The results for the other simulation models can be explained by Theorem 1 and Corollary 1 in a similar way. For example, under the simulation model $(\alpha_{i\cdot}, \beta_{\cdot j})$, Theorem 1-2) and 4) show that $\nu(j,j') \in OI^{\nu}(j,j')$ and $\omega(i,i') \in OI^{\omega}(i,i')$ always hold, independent of the magnitude of $\lambda^{11}_{Y_1R_1}$ and $\lambda^{11}_{Y_2R_2}$, respectively. This means that neither C_1 nor C_2 in Corollary 1 hold and all four saturated models are plausible. For the simulation model $(\alpha_{\cdot j}, \beta_{i\cdot})$, the occurrence of $\nu(j,j') \notin OI^{\nu}(j,j')$ and $\omega(i,i') \notin OI^{\omega}(i,i')$ depend on the magnitude of $\lambda^{11}_{Y_2R_1}$ and $\lambda^{11}_{Y_1R_2}$, respectively, by Theorem 1-1) and 3). Thus, when the values of $\lambda^{11}_{Y_2R_1}$ and $\lambda^{11}_{Y_1R_2}$ are both large, both conditions C_1 and C_2 in Corollary 1 hold, so the selected model is only $(\alpha_{\cdot j}, \beta_{i\cdot})$.

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0472

Table 4: Number of the cases where each of the four saturated models was chosen by AIC, BIC, G^2 and the proposed method Eq. (4.4) in Corollary 1 under the four simulation models (types of nonignorable missingness) with N=5000.

Simulation model $(\alpha_{i\cdot}, \beta_{i\cdot})$		Fitted saturated models				Simulation model $(\alpha_{\cdot j}, \beta_{\cdot j})$		Fitted saturated models			
$\lambda^{11}_{Y_1R_1}$	$\lambda^{11}_{Y_1R_2}$	$(\alpha_{i\cdot}, \beta_{i\cdot})$	$(\alpha_{i\cdot}, \beta_{\cdot j})$	$(\alpha_{\cdot j}, \beta_{i\cdot})$	$(\alpha_{\cdot j}, \beta_{\cdot j})$	$\lambda^{11}_{Y_2R_1}$	$\lambda^{11}_{Y_2R_2}$	$(\alpha_{i\cdot}, \beta_{i\cdot})$	$(\alpha_{i\cdot}, \beta_{\cdot j})$	$(\alpha_{\cdot j}, \beta_{i\cdot})$	$(\alpha_{\cdot j}, \beta_{\cdot j})$
0.05	0.05	1000	996	1000	996	0.05	0.05	998	998	1000	1000
0.1	0.05	1000	992	1000	992	0.1	0.05	803	803	1000	1000
0.2	0.05	1000	992	1000	992	0.2	0.05	1	1	1000	1000
0.4	0.05	1000	992	1000	992	0.4	0.05	0	0	1000	1000
0.05	0.1	1000	634	1000	634	0.05	0.1	996	996	1000	1000
0.1	0.1	1000	632	1000	632	0.1	0.1	768	768	1000	1000
0.2	0.1	1000	644	1000	644	0.2	0.1	1	1	1000	1000
0.4	0.1	1000	652	1000	652	0.4	0.1	0	0	1000	1000
0.05	0.2	1000	0	1000	0	0.05	0.2	998	998	1000	1000
0.1	0.2	1000	0	1000	0	0.1	0.2	789	789	1000	1000
0.2	0.2	1000	0	1000	0	0.2	0.2	0	0	1000	1000
0.4	0.2	1000	0	1000	0	0.4	0.2	0	0	1000	1000
0.05	0.4	1000	0	1000	0	0.05	0.4	1000	1000	1000	1000
0.1	0.4	1000	0	1000	0	0.1	0.4	769	769	1000	1000
0.2	0.4	1000	0	1000	0	0.2	0.4	1	1	1000	1000
0.4	0.4	1000	0	1000	0	0.4	0.4	0	0	1000	1000

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0472

Simulation model $(\alpha_{i\cdot}, \beta_{\cdot j})$		Fitted saturated models				Simulation model $(\alpha_{\cdot j}, \beta_{i\cdot})$		Fitted saturated models			
$\lambda^{11}_{Y_1R_1}$	$\lambda^{11}_{Y_2R_2}$	$(\alpha_{i\cdot}, \beta_{i\cdot})$	$(\alpha_{i\cdot}, \beta_{\cdot j})$	$(\alpha_{\cdot j}, \beta_{i\cdot})$	$(\alpha_{\cdot j}, \beta_{\cdot j})$	$\lambda^{11}_{Y_2R_1}$	$\lambda^{11}_{Y_1R_2}$	$(\alpha_{i\cdot}, \beta_{i\cdot})$	$(\alpha_{i\cdot}, \beta_{\cdot j})$	$(\alpha_{\cdot j}, \beta_{i\cdot})$	$(\alpha_{\cdot j}, \beta_{\cdot j})$
0.05	0.05	1000	1000	1000	1000	0.05	0.05	998	988	1000	990
0.1	0.05	1000	1000	1000	1000	0.1	0.05	774	768	1000	992
0.2	0.05	1000	1000	1000	1000	0.2	0.05	1	1	1000	990
0.4	0.05	1000	1000	1000	1000	0.4	0.05	0	0	1000	995
0.05	0.1	1000	1000	1000	1000	0.05	0.1	996	635	1000	638
0.1	0.1	1000	1000	1000	1000	0.1	0.1	786	531	1000	666
0.2	0.1	1000	1000	1000	1000	0.2	0.1	1	1	1000	663
0.4	0.1	1000	1000	1000	1000	0.4	0.1	0	0	1000	639
0.05	0.2	1000	1000	1000	1000	0.05	0.2	998	1	1000	1
0.1	0.2	1000	1000	1000	1000	0.1	0.2	784	0	1000	0
0.2	0.2	1000	1000	1000	1000	0.2	0.2	0	0	1000	0
0.4	0.2	1000	1000	1000	1000	0.4	0.2	0	0	1000	1
0.05	0.4	1000	1000	1000	1000	0.05	0.4	997	0	1000	0
0.1	0.4	1000	1000	1000	1000	0.1	0.4	784	0	1000	0
0.2	0.4	1000	1000	1000	1000	0.2	0.4	2	0	1000	0
0.4	0.4	1000	1000	1000	1000	0.4	0.4	0	0	1000	0

5.2 Data analysis

Table 5 shows an incomplete 3×3 table, from the third National Health and Nutrition Examination Survey, classified by bone mineral density (BMD, Y_1) and family income (FI, Y_2) (Nandram, Cox, and Choi (2005)). Using Table 5 we illustrate how the proposed methods, (4.4) and (4.5), can be used in assessing the plausibility of nonignorable log-linear models in Table 2. We also consider the ignorable nonresponse models as the important baseline models, MCAR model ($\alpha_{\cdot\cdot}$, $\beta_{\cdot\cdot}$) in Remark 1, and the MAR selection model in Molenberghs, Beunckens, Sotto, and Kenward (2008). For the details of the MAR selection model, see Section S2.2 in the

Table 5: BMD (Y_1) and FI (Y_2) data

		$R_2 = 1$			$R_2 = 2$
		$Y_2 = 1$ (< \$20,000)	$Y_2 = 2$ (\$20,000, \$45,000)	$Y_2 = 3$ (>\$45,000)	Y_2
	$Y_1 = 1$ (normal)	621	290	284	135
$R_1 = 1$	$Y_1 = 2$ (osteopenia)	260	131	117	69
	$Y_1 = 3$ (osteoporosis)	93	30	18	27
$R_1 = 2$	Y_1	456	156	266	45

Table 6: $\hat{\nu}_i(j, j')$ and $\hat{\nu}(j, j')$ for the models $(\alpha_{i\cdot},\ \beta_{\square})$ and $(\alpha_{\cdot j},\ \beta_{\square})$, and $\hat{\omega}_j(i, i')$ and $\hat{\omega}(i, i')$ for the models $(\alpha_{\square},\ \beta_{i\cdot})$ and $(\alpha_{\square},\ \beta_{\cdot j})$

(j, j')	$\nu_1(j, j')$	$\nu_2(j, j')$	$\nu_3(j, j')$	$\nu(j, j')$
(1,2)	2.14(=621/290)	1.98(=260/131)	3.10(=93/30)	2.92(=456/156)
(1,3)	2.19(=621/284)	2.22(=260/117)	5.17(=93/18)	1.71(=456/266)
(2,3)	1.02(=290/284)	1.12(=131/117)	1.67(=30/18)	0.59(=156/266)

(i, i')	$\omega_1(i, i')$	$\omega_2(i, i')$	$\omega_3(i, i')$	$\omega(i, i')$
(1,2)	2.39(=621/260)	2.21(=290/131)	2.43(=284/117)	1.96(=135/69)
(1,3)	6.68(=621/93)	9.67(=290/30)	15.78(=284/18)	5.00(=135/27)
(2,3)	2.80(=260/93)	4.37(=131/30)	6.50(=117/18)	2.56(=69/27)

Supplementary Material.

We evaluate the conditions C_1 and C_2 of Corollary 1 by using the estimators in Table 3 to compute $\hat{\nu}(j, j')$ in (4.1), $\hat{\omega}(i, i')$ in (4.2), and $\widehat{OI}^{\nu}(j, j')$ and $\widehat{OI}^{\omega}(i, i')$ in (4.3) under the saturated models $(\alpha_{i\cdot},\ \beta_{i\cdot})$, $(\alpha_{\cdot j},\ \beta_{\cdot j})$, $(\alpha_{i\cdot},\ \beta_{\cdot j})$, and $(\alpha_{\cdot j},\ \beta_{i\cdot})$.

From Table 6 we observe that both conditions C_1 and C_2 hold: $\hat{\nu}(j, j') \notin \widehat{OI}^{\nu}(j, j')$ for two pairs (j, j') of Y_2 and $\hat{\omega}(i, i') \notin \widehat{OI}^{\omega}(i, i')$ for three pairs

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0472

(i, i') of Y_1. Thus, by (4.4), the plausible saturated nonignorable log-linear model is only $(\alpha_{\cdot j},\ \beta_{i\cdot})$, the model with $\lambda^{jk}_{Y_2R_1}$ and $\lambda^{i\ell}_{Y_1R_2}$. To assess the uncertainty of the accuracy of the proposed method, we performed bootstrap resampling: generate 100,000 samples from the model $(\alpha_{\cdot j},\ \beta_{i\cdot})$ fitted to the data and compute the percentages of bootstrap samples satisfying the conditions C_1 and C_2 in Corollary 1. The computed percentages for C_1 and C_2 were 99.99 and 96.47, respectively, which confirm the accuracy of the proposed method.

We also compared the result of the proposed method with the model selection results of G^2, AIC, and BIC in selecting the suitable saturated models. As shown in Table 7, G^2, AIC, and BIC chose the same saturated model as the proposed method, $(\alpha_{\cdot j}, \beta_{i\cdot})$: the model $(\alpha_{\cdot j}, \beta_{i\cdot})$ has a zero value of G^2 and the smallest values of AIC and BIC among all saturated models. The expected cell counts given in Section S5.1 of the Supplementary Material shows that the other three saturated models gave a poor fit to the observed cell counts and/or the boundary solutions for the estimates of the unobserved cell counts.

For the statistical significance of the interaction parameters in the selected model $(\alpha_{\cdot j}, \beta_{i\cdot})$, as suggested in (4.5), we considered the corresponding nested models $(\alpha_{\cdot\cdot},\ \beta_{i\cdot})$ and $(\alpha_{\cdot j},\ \beta_{\cdot\cdot})$ and used the LRT, AIC, BIC,

Table 7: Model selection for BMD data. Note that $(C_1,\ C_2)$ represents the percentage of bootstrap samples satisfying the conditions C_1 and C_2 in Corollary 1

Saturated model	Nested model	Proposed method (C_1, C_2)	P-value for LRT	G^2	AIC	BIC
$(\alpha_{i\cdot}, \beta_{\cdot j})$				20.6	14,574.51	14,585.13
	$(\alpha_{i\cdot}, \beta_{\cdot\cdot})$		0.219	23.6	14,573.55	14,582.75
	$(\alpha_{\cdot\cdot}, \beta_{\cdot j})$		0.024	28.0	14,577.96	14,587.16
$(\alpha_{\cdot j}, \beta_{i\cdot})$		✓(99.99, 96.47)		**0**	**14,553.93**	**14,564.55**
	$(\alpha_{\cdot j}, \beta_{\cdot\cdot})$	✓	0.066	5.4	**14,555.36**	**14,564.56**
	$(\alpha_{\cdot\cdot}, \beta_{i\cdot})$	✓	< 0.001	25.7	14,575.59	14,584.79
$(\alpha_{i\cdot}, \beta_{i\cdot})$				18.2	14,572.12	14,582.74
	$(\alpha_{i\cdot}, \beta_{\cdot\cdot})$		0.066	23.6	14,573.55	14,582.75
	$(\alpha_{\cdot\cdot}, \beta_{i\cdot})$		0.024	25.7	14,575.59	14,584.79
$(\alpha_{\cdot j}, \beta_{\cdot j})$				2.4	14,556.29	14,566.91
	$(\alpha_{\cdot j}, \beta_{\cdot\cdot})$		0.215	5.4	14,555.36	14,564.56
	$(\alpha_{\cdot\cdot}, \beta_{\cdot j})$		< 0.001	28.0	14,577.96	14,587.16

and G^2. From Table 7, the nested model $(\alpha_{\cdot j},\ \beta_{\cdot\cdot})$ must also be selected, as the p-values of LRT is 0.066 and the AIC and BIC values between the full model $(\alpha_{\cdot j}, \beta_{i\cdot})$ and the reduced model $(\alpha_{\cdot j},\ \beta_{\cdot\cdot})$ have only small differences.

Finally, we applied the ignorable nonresponse models to the data, the MCAR model $(\alpha_{\cdot\cdot}, \beta_{\cdot\cdot})$ and the MAR selection model. The values of G^2, AIC, and BIC were (31.27, 14,577.20, 14,584.99) for $(\alpha_{\cdot\cdot}, \beta_{\cdot\cdot})$ and (0, 14,553.93, 14,564.55) for the MAR model. The MAR model performs better than the

MCAR model. The estimates for the expected cell counts under the ignorable nonresponse models are given in Section S5.1 of the Supplementary Material.

From this analysis, we conclude that the plausible nonignorable log-linear models for Table 5 are $(\alpha_{.j}, \beta_{i.})$ and $(\alpha_{.j},\ \beta_{..})$, and the MAR selection model has the same results as the selected saturated model $(\alpha_{.j}, \beta_{i.})$ in terms of G^2, AIC, BIC and the estimates for the observed data, $\hat{m}_{ij11}$. However, the MAR model and the selected saturated models $(\alpha_{.j}, \beta_{i.})$ produce different predictions of the unobserved cell counts. This means that the MAR selection model is the MAR counterpart of the saturated model $(\alpha_{.j}, \beta_{i.})$, and thus the empirical distinction between MAR and MNAR is not possible (Molenberghs, Beunckens, Sotto, and Kenward, 2008).

6. Discussion

We have proposed a data analytic method that aims to select nonignorable log-linear models suitable for an incomplete two-way contingency table. The proposed method involves only the computation of a set of response odds and nonresponse odds that can be easily obtained from observed data. The simulation results and data analysis showed the same performance between the proposed method and the standard model selection criteria, AIC, BIC, and G^2 in selecting plausible saturated models. For the nonignorable log-linear model(s) selected by the proposed method, one can perform a sensitivity analysis to better understand the parameters of interest (Molenberghs, Kenward, and Goetghebeur (2001); Vansteelandt, Goetghebeur, Kenward, and Molenberghs (2006)).

We here discuss the generalizability of the proposed method in two directions : its applicability for a more than two-way incomplete contingency table and its applicability for nonignorable models outside the framework of log-linear models.

We considered hierarchical log-linear models for three-way incomplete contingency tables with the variables subject to missingness. As shown in Section S7 of the Supplemental Material, we confirmed that the proposed method can be easily generalized to identify the types of misssingness assumed in the log-linear models for incomplete three-way tables; the inequalities relating the nonresponse odds to the response odds intervals enable one to identify the informative missingness represented by the (two-way and three-way) interaction parameters of the assumed log-linear model.

We examined the applicability of the proposed method to a nonignorable selection model (Fay (1986); Molenberghs, Goetghebeur, Lipsitz, and Kenward (1999)) for an incomplete two-way table. As given in Section S8 of the Supplemental Material, the inequalities relating the nonresponse odds to the response odds intervals are directly associated with the magnitudes of the parameters describing missingness assumed in the selection model. We also demonstrated their performance on data.

Appendix A.

A.1 Proof of Theorem 1 in Section 3

We show only the proof of Theorem 1-1) and 2). The proofs of Theorem 1-3) and 4) are provided in Section S6 of the Supplementary Material.

Lemma 1. *For a $I \times I \times 2 \times 2$ contingency table, the following inequalities hold under all nonresponse models in Table 2. For all pairs (j, j') of Y_2,*

$$\lambda^{ij'}_{Y_1Y_2} - \lambda^{mj'}_{Y_1Y_2} > \lambda^{ij}_{Y_1Y_2} - \lambda^{mj}_{Y_1Y_2} \quad \textit{and} \quad \lambda^{ij'}_{Y_1Y_2} - \lambda^{nj'}_{Y_1Y_2} < \lambda^{ij}_{Y_1Y_2} - \lambda^{nj}_{Y_1Y_2}$$

where m and n are subscripts corresponding to $\nu_m(j, j')$ and $\nu_n(j, j')$, respectively.

Proof. We show the proof for the model $(\alpha_{i\cdot}, \beta_{\cdot j})$. The proofs for the other models are similar. First, $\nu_i(j, j')$ is expressed as

$$\nu_i(j, j') = \frac{\pi_{ij11}}{\pi_{ij'11}} = \exp\left(\lambda^{j}_{Y_2} - \lambda^{j'}_{Y_2} + \lambda^{ij}_{Y_1Y_2} - \lambda^{ij'}_{Y_1Y_2} + \lambda^{j1}_{Y_2R_2} - \lambda^{j'1}_{Y_2R_2}\right) \quad (4.6)$$

for each pair (j, j') of Y_2. Comparison of $\nu_i(j, j')$ and $\nu_m(j, j')$ in terms of (4.6) gives $\lambda^{ij'}_{Y_1Y_2} - \lambda^{mj'}_{Y_1Y_2} > \lambda^{ij}_{Y_1Y_2} - \lambda^{mj}_{Y_1Y_2}$, because $\nu_m(j, j') \geq \nu_i(j, j')$ for all i

Table 8: Decomposition of $R^{\nu}_{m}(j,j')$ and $R^{\nu}_{n}(j,j')$

Model	$H^{\nu}(j,j')$	$M^{\nu}_{m}(j,j')$	$M^{\nu}_{n}(j,j')$
$(\alpha_{\cdot j},\beta_{\cdot j})$ $(\alpha_{\cdot j},\beta_{\cdot\cdot})$	$\exp\left(2\lambda^{j'2}_{Y_2R_1}-2\lambda^{j2}_{Y_2R_1}\right)$	$\frac{\sum_i \exp\left(\lambda^{i}_{Y_1}+\lambda^{ij'}_{Y_1Y_2}-\lambda^{mj'}_{Y_1Y_2}\right)}{\sum_i \exp\left(\lambda^{i}_{Y_1}+\lambda^{ij}_{Y_1Y_2}-\lambda^{mj}_{Y_1Y_2}\right)}$	$\frac{\sum_i \exp\left(\lambda^{i}_{Y_1}+\lambda^{ij'}_{Y_1Y_2}-\lambda^{nj'}_{Y_1Y_2}\right)}{\sum_i \exp\left(\lambda^{i}_{Y_1}+\lambda^{ij}_{Y_1Y_2}-\lambda^{nj}_{Y_1Y_2}\right)}$
$(\alpha_{\cdot j},\beta_{i\cdot})$	$\exp\left(2\lambda^{j'2}_{Y_2R_1}-2\lambda^{j2}_{Y_2R_1}\right)$	$\frac{\sum_i \exp\left(\lambda^{i}_{Y_1}+\lambda^{i1}_{Y_1R_2}+\lambda^{ij'}_{Y_1Y_2}-\lambda^{mj'}_{Y_1Y_2}\right)}{\sum_i \exp\left(\lambda^{i}_{Y_1}+\lambda^{i1}_{Y_1R_2}+\lambda^{ij}_{Y_1Y_2}-\lambda^{mj}_{Y_1Y_2}\right)}$	$\frac{\sum_i \exp\left(\lambda^{i}_{Y_1}+\lambda^{i1}_{Y_1R_2}+\lambda^{ij'}_{Y_1Y_2}-\lambda^{nj'}_{Y_1Y_2}\right)}{\sum_i \exp\left(\lambda^{i}_{Y_1}+\lambda^{i1}_{Y_1R_2}+\lambda^{ij}_{Y_1Y_2}-\lambda^{nj}_{Y_1Y_2}\right)}$
$(\alpha_{\cdot\cdot},\beta_{\cdot j})$	1	$\frac{\sum_i \exp\left(\lambda^{i}_{Y_1}+\lambda^{ij'}_{Y_1Y_2}-\lambda^{mj'}_{Y_1Y_2}\right)}{\sum_i \exp\left(\lambda^{i}_{Y_1}+\lambda^{ij}_{Y_1Y_2}-\lambda^{mj}_{Y_1Y_2}\right)}$	$\frac{\sum_i \exp\left(\lambda^{i}_{Y_1}+\lambda^{ij'}_{Y_1Y_2}-\lambda^{nj'}_{Y_1Y_2}\right)}{\sum_i \exp\left(\lambda^{i}_{Y_1}+\lambda^{ij}_{Y_1Y_2}-\lambda^{nj}_{Y_1Y_2}\right)}$
$(\alpha_{\cdot\cdot},\beta_{i\cdot})$	1	$\frac{\sum_i \exp\left(\lambda^{i}_{Y_1}+\lambda^{i1}_{Y_1R_2}+\lambda^{ij'}_{Y_1Y_2}-\lambda^{mj'}_{Y_1Y_2}\right)}{\sum_i \exp\left(\lambda^{i}_{Y_1}+\lambda^{i1}_{Y_1R_2}+\lambda^{ij}_{Y_1Y_2}-\lambda^{mj}_{Y_1Y_2}\right)}$	$\frac{\sum_i \exp\left(\lambda^{i}_{Y_1}+\lambda^{i1}_{Y_1R_2}+\lambda^{ij'}_{Y_1Y_2}-\lambda^{nj'}_{Y_1Y_2}\right)}{\sum_i \exp\left(\lambda^{i}_{Y_1}+\lambda^{i1}_{Y_1R_2}+\lambda^{ij}_{Y_1Y_2}-\lambda^{nj}_{Y_1Y_2}\right)}$
$(\alpha_{i\cdot},\beta_{\cdot j})$ $(\alpha_{i\cdot},\beta_{\cdot\cdot})$	1	$\frac{\sum_i \exp\left(\lambda^{i}_{Y_1}+\lambda^{i2}_{Y_1R_1}+\lambda^{ij'}_{Y_1Y_2}-\lambda^{mj'}_{Y_1Y_2}\right)}{\sum_i \exp\left(\lambda^{i}_{Y_1}+\lambda^{i2}_{Y_1R_1}+\lambda^{ij}_{Y_1Y_2}-\lambda^{mj}_{Y_1Y_2}\right)}$	$\frac{\sum_i \exp\left(\lambda^{i}_{Y_1}+\lambda^{i2}_{Y_1R_1}+\lambda^{ij'}_{Y_1Y_2}-\lambda^{nj'}_{Y_1Y_2}\right)}{\sum_i \exp\left(\lambda^{i}_{Y_1}+\lambda^{i2}_{Y_1R_1}+\lambda^{ij}_{Y_1Y_2}-\lambda^{nj}_{Y_1Y_2}\right)}$
$(\alpha_{i\cdot},\beta_{i\cdot})$	1	$\frac{\sum_i \exp\left(\lambda^{i}_{Y_1}+\lambda^{i2}_{Y_1R_1}+\lambda^{i1}_{Y_1R_2}+\lambda^{ij'}_{Y_1Y_2}-\lambda^{mj'}_{Y_1Y_2}\right)}{\sum_i \exp\left(\lambda^{i}_{Y_1}+\lambda^{i2}_{Y_1R_1}+\lambda^{i1}_{Y_1R_2}+\lambda^{ij}_{Y_1Y_2}-\lambda^{mj}_{Y_1Y_2}\right)}$	$\frac{\sum_i \exp\left(\lambda^{i}_{Y_1}+\lambda^{i2}_{Y_1R_1}+\lambda^{i1}_{Y_1R_2}+\lambda^{ij'}_{Y_1Y_2}-\lambda^{nj'}_{Y_1Y_2}\right)}{\sum_i \exp\left(\lambda^{i}_{Y_1}+\lambda^{i2}_{Y_1R_1}+\lambda^{i1}_{Y_1R_2}+\lambda^{ij}_{Y_1Y_2}-\lambda^{nj}_{Y_1Y_2}\right)}$

by the definition of $\nu_m(j,j')$. Similarly, we have $\lambda^{ij'}_{Y_1Y_2}-\lambda^{nj'}_{Y_1Y_2} < \lambda^{ij}_{Y_1Y_2}-\lambda^{nj}_{Y_1Y_2}$ because $\nu_n(j,j') \leq \nu_i(j,j')$ for all i by the definition of $\nu_n(j,j')$. $\square$

We turn to proofs of Theorem 1-1) and 2).

Proof. For all models given in Table 2, let $R^{\nu}_{m}(j,j') = \nu_m(j,j')/\nu(j,j')$ and $R^{\nu}_{n}(j,j') = \nu_n(j,j')/\nu(j,j')$. Then, $R^{\nu}_{m}(j,j')$ and $R^{\nu}_{n}(j,j')$ are decomposed into two parts according to each model as shown in Table 8 : $R^{\nu}_{m}(j,j') = H^{\nu}(j,j')M^{\nu}_{m}(j,j')$ and $R^{\nu}_{n}(j,j') = H^{\nu}(j,j')M^{\nu}_{n}(j,j')$. By Lemma 1, for all eight log-linear models, we have $M^{\nu}_{m}(j,j') > 1$ for all pairs (j,j') of Y_2, as all other parameters of $M^{\nu}_{m}(j,j')$ are same except $\lambda^{ij'}_{Y_1Y_2}-\lambda^{mj'}_{Y_1Y_2}$ in the numerator and $\lambda^{ij}_{Y_1Y_2}-\lambda^{mj}_{Y_1Y_2}$ in the denominator, as shown in Table 8. On the contrary, by Lemma 1, we have $M^{\nu}_{n}(j,j') < 1$ for all pairs (j,j') of

Y_2 under all eight models. We have $M_m^\nu(j,j')$ and $M_n^\nu(j,j')$ that equal to $R_m^\nu(j,j')$ and $R_n^\nu(j,j')$, respectively, under the assumption $\lambda_{Y_2R_1}^{j2}=0$ for all j's (which makes $H^\nu(j,j')=1$ for all j's).

Under the models $(\alpha_{\cdot j},\beta_\square)$ (i.e., $(\alpha_{\cdot j},\beta_{\cdot j})$, $(\alpha_{\cdot j},\beta_{i\cdot})$, and $(\alpha_{\cdot j},\beta_{\cdot\cdot})$), we show the necessary and sufficient condition for $\nu(j,j')\in OI^\nu(j,j')$. First, $\nu(j,j')\in OI^\nu(j,j')$ is the same as "$R_m^\nu(j,j')>1$ and $R_n^\nu(j,j')<1$". Since $R_m^\nu(j,j')=H^\nu(j,j')M_m^\nu(j,j')$ and $R_n^\nu(j,j')=H^\nu(j,j')M_n^\nu(j,j')$, "$R_m^\nu(j,j')>1$ and $R_n^\nu(j,j')<1$" is equivalent to that $H^\nu(j,j')$ is larger than $M_m^\nu(j,j')^{-1}$ and less than $M_n^\nu(j,j')^{-1}$, and thus $-\log M_m^\nu(j,j')<2(\lambda_{Y_2R_1}^{j'2}-\lambda_{Y_2R_1}^{j2})=\log H^\nu(j,j')<-\log M_n^\nu(j,j')$. Since $\nu(j,j')\notin OI^\nu(j,j')$ is the complement of $\nu(j,j')\in OI^\nu(j,j')$, it is straightforward to show that the necessary and sufficient condition for $\nu(j,j')\notin OI^\nu(j,j')$ is "$\log H^\nu(j,j')=2(\lambda_{Y_2R_1}^{j'2}-\lambda_{Y_2R_1}^{j2})<-\log M_m^\nu(j,j')$ or $\log H^\nu(j,j')=2(\lambda_{Y_2R_1}^{j'2}-\lambda_{Y_2R_1}^{j2})>-\log M_n^\nu(j,j')$".

For the models $(\alpha_{i\cdot},\beta_\square)$ (i.e., $(\alpha_{i\cdot},\beta_{\cdot j})$, $(\alpha_{i\cdot},\beta_{i\cdot})$, and $(\alpha_{i\cdot},\beta_{\cdot\cdot})$), $R_m^\nu(j,j')=M_m^\nu(j,j')$ and $R_n^\nu(j,j')=M_n^\nu(j,j')$, because $H^\nu(j,j')=1$ for all pairs (j,j') of Y_2 as shown in Table 8. Since $M_m^\nu(j,j')>1$ and $M_m^\nu(j,j')<1$, we have $\nu(j,j')\in OI^\nu(j,j')$ for all pairs (j,j') of Y_2. For the models $(\alpha_{\cdot\cdot},\beta_{i\cdot})$ and $(\alpha_{\cdot\cdot},\beta_{\cdot j})$, $H^\nu(j,j')$=1 for all pairs (j,j') of Y_2. So, $R_m^\nu(j,j')$=$M_m^\nu(j,j')$ and $R_n^\nu(j,j')$=$M_n^\nu(j,j')$, which means $\nu(j,j')\in OI^\nu(j,j')$. □

doi:10.5705/ss.202015.0472

Supplementary Materials

In the Supplemental Material, we provide some details on the eight log-linear models given in Table 2, on the ignorable nonresponse models for an incomplete $I \times J$ table, on the results of simulation studies and data analysis that were omitted in Section 5, on the proofs of Theorem 1-3) and 4) in the paper, on the log-linear models for incomplete three-way contingency tables, and on the nonignorable selection model for an incomplete two-way contingency table.

Acknowledgements

The authors thank the Editor, the associate editor, and the referee for the comments that substantially improved the article. This research was supported by Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education (NRF-2016R1D1A3B03930392).

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**, 716-723.

Baker, S. G., Ko, C., and Graubard, B. I. (2003). A sensitivity analysis for nonrandomly missing categorical data arising from a national health disability survey. *Biostatistics* **4**, 41-56.

Baker, S. G., and Laird, N. M. (1988). Regression analysis for categorical variables with outcome subject to nonignorable nonresponse. *J. Amer. Statist. Assoc.* **83**, 62-69.

Baker, S. G., Rosenberger, W. F., and Dersimonian, R. (1992). Closed-form estimates for missing counts in two-way contingency tables. *Stat. Med.* **11**, 643-657.

Clarke, P. S. and Smith, P. W. F. (2004). Interval estimation for log-linear models with one variable subject to non-ignorable non-response. *J. Roy. Statist. Soc. Ser. B* **66**, 357-368.

Conaway, M. R., Waternaux, C., Allred, E., Bellinger, D., and Leviton, A. (1992). Pre-natal blood lead levels and learning difficulties in children: An analysis of non-randomly missing categorical data. *Stat. Med.* **11**, 799-811.

Copas, J. B. and Eguchi, S. (2001). Local sensitivity approximation for selectivity bias. *J. Roy. Statist. Soc. Ser. B*, **63**, 871–895.

Efron, B. and Tibshirani R. J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall, New York.

Fay, R. E. (1986). Causal Models for Patterns of Nonresponse. *J. Amer. Statist. Assoc.* **81**, 354-365.

Forster, J. J. and Smith, P. W. F. (1998). Model-based inference for categorical survey data subject to nonignorable nonresponse. *J. Roy. Statist. Soc. Ser. B* **60**, 57-70.

Little, R. J. A. (1993). Pattern-mixture models for multivariate incomplete data. *J. Amer.*

Statist. Assoc. **88**, 125-134.

Little, J.A. and Rubin, D.B. (2002). *Statistical Analysis with Missing Data, 2nd ed.* Wiley, New York.

Molenberghs, G., Beunckens, C., Sotto, C., and Kenward, M. G. (2008). Every missingness not at random model has a missingness at random counterpart with equal fit. *J. Roy. Statist. Soc. Ser. B* **70**, 371-388.

Molenberghs, G., Goetghebeur, E. J. T., Lipsitz, S. R., and Kenward, M. G. (1999). Non-random missingness in categorical data: Strengths and limitations. *Am. Stat.* **53**, 110-118.

Molenberghs, G., Kenward, M.G., and Goetghebeur, E. (2001). Sensitivity analysis for incomplete contingency tables: the Slovenian plebiscite case. *J. Roy. Statist. Soc. Ser. C* **50**, 15-29.

Nandram, B., Cox, L. H., and Choi, J. W. (2005). Bayesian Analysis of Nonignorable Missing Categorical Data: An Application to Bone Mineral Density and Family Income. *Surv. Methodol.* **31** : 213-225.

Park, T. and Brown, M. B. (1994). Models for categorical data with nonignorable nonresponse. *J. Amer. Statist. Assoc. 89*, 44-52.

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

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* **6**, 461-464.

Troxel, A.B., Ma, G., and Heitjan, D. F. (2004). An index of local sensitivity to nonignorability.

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0472

Statist. Sinica **6**, 1221–1237 .

Vansteelandt, S., Goetghebeur, E., Kenward, M. G., and Molenberghs, G. (2006). Ignorance and uncertainty regions as inferential tools in a sensitivity analysis. *Statist. Sinica* **16**, 953-979.

Xie, H., Qian, Y., and Qu, L. (2011). A semiparametric approach for analyzing nonignorable missing data. *Statist. Sinica* **21**, 1881–1899.

Seongyong Kim, Department of Applied Statistics, Hoseo University, Asan, South Korea

Daeyoung Kim, Department of Mathematics & Statistics, University of Massachusetts Amherst, US.

E-mail: daeyoung@math.umass.edu