

文章编号: 1000-5862(2017)03-0302-06

缺失数据比率和处理方法对非随机缺失 数据能力参数估计准确性的影响

康春花, 孙金玲, 孙小坚, 曾平飞*

(浙江师范大学教师教育学院, 浙江 金华 321004)

摘要: 探讨了 IRT 背景下非随机缺失数据的合适处理方法. 采用 IRTLAB 模拟产生 50 批 500 个被试在 20 个 0-1 记分项目上的反应数据, 产生了不同比率的 MNAR; 再用 IN、NP、FR、CM、MI 和 EM 共 6 种方法分别处理 MNAR, 使用 BILOG-MG 软件估计被试的能力参数, 并计算在不同条件下各种方法的 $BIAS$ 、 $BIAS_{abs}$ 、 $R_{(\theta, \hat{\theta})}$ 和 $RMSE$. 研究发现: 随着缺失比率的增加, 参数误差越来越大; FR 会导致 IRT 参数估计产生较大的误差, 且不稳定, 而 MI 与 EM 算法则相对稳定; 综合 $BIAS$ 和 $RMAE$ 等几个指标, NP 在处理 MNAR 时产生的误差较小也更稳定. 因此, 在 IRT 背景下估计被试能力参数时, 应选择 NP、MI 或 EM 方法处理缺失数据.

关键词: 缺失数据比例; 缺失数据处理方法; IRT 参数估计

中图分类号: B 841.7 **文献标志码:** A **DOI:** 10.16357/j.cnki.issn1000-5862.2017.03.17

0 引言

在社会、教育和心理领域的调查和实验研究中, 常常会出现数据缺失的情形. 缺失数据是指未能取得所要收集的研究数据, 是由一个或多个被试在一个或多个研究变量上未反应引起的. 一般而言, 缺失数据的产生原因主要有以下几种: 1) 测验设计, 如国际大型学业测评项目 PISA 采用矩阵取样的设计技术, 测验中使用多个题本, 但每个被试只做其中一个题本, 这样被试在其它题本上的作答就是缺失的; 2) 由于时间有限, 导致被试没有机会作答完所有的项目; 3) 被试能力不足或缺乏动机从而放弃对一些项目的作答^[1-2]. 通常, 根据缺失数据产生的原因或存在的形式, 可将其分为完全随机缺失数据 (missing completely at random, MCAR)、随机缺失数据 (missing at random, MAR) 和非随机缺失数据 (missing not at random, MNAR)^[3-4]. 在一个作答矩阵中, 缺失数据用 Y_{mis} 来表示, 可观测数据用 Y_{obs} 来表示, 则整个作答矩阵可以表示为 $Y = [Y_{mis}, Y_{obs}]$. 根据 Little 的定义, MCAR 的产生既不依赖于 Y_{mis} 也不依赖于 Y_{obs} , 例如由于矩阵取样导致的缺失数据即为 MCAR. 把 MCAR 的约束进一步放宽, 当缺失数据的产生依赖于 Y_{obs} 就得到 MAR, 例如, 由于被试生病

而不能参加测验导致的缺失数据. 当随机缺失数据的前提假设不成立时, 即缺失数据的产生依赖于 Y_{mis} 或其他不可观测的协变量时, 就产生了 MNAR, 例如当被试由于能力不足不能对题目作答而导致的缺失数据. 在社会、教育、心理及其它领域, 由于不完整的反应数据或缺失数据的存在, 会影响人们对被试真实能力的评判, 从而导致参数的有偏估计、统计检验力的降低, 最终影响着研究的信度和测评的公平. 因此, 从统计和测量的角度, 识别导致缺失数据产生的变量和机制, 探讨合适的缺失数据处理方法就显得尤为重要^[5-11].

Holmes Finch^[12] 研究了在 MNAR 和 MAR 的 2 种缺失数据类型下 7 种缺失数据处理方法对 IRT 项目参数估计准确性的影响, 发现错误作答 (incorrect IN) 是参数估计准确性最差的方法, 相比之下, 忽略不计 (not presented, NP) 的误差较小; 对于 MNAR 而言, 除 IN 方法外, 其它方法倾向于低估难度参数, 其中 EM 方法对难度参数低估的程度最高^[12]. Norman 在国际大型学业监测项目 PISA 中, 考查了几种基于模型的 MNAR 处理方法对参数估计准确性的影响, 研究结果表明参数估计偏差与潜在特质、反应倾向的关联程度和缺失数据比率有关. 在中等缺失比率下 (30%) 项目参数和被试能力参数相当稳健, 准确性也较高; 在高缺失比率下

收稿日期: 2016-09-22

基金项目: 浙江省自然科学基金 (LY15C090003) 和教育部人文社会科学基金 (16YJA190002) 资助项目.

通信作者: 曾平飞 (1963-), 男, 广西荔浦人, 教授, 博士, 主要从事心理测量与评价方面的研究. E-mail: zpf@zjnu.edu.cn

(50%) ,项目参数和被试能力参数均会受到严重影响; 错误作答(IN) 似乎是最差的 MNAR 处理方法, 基于模型的方法则比较合适^[6]. 国内关于 IRT 缺失数据的研究起步较晚, 数量也较少. 张淑梅等^[13] 采用模拟研究考察了 EM 算法与 IN 方法对随机缺失数据 IRT 项目参数估计准确性的影响. 李斌等^[9] 基于潜变量建模, 采用 GIBBS 抽样对 MNAR 参数估计进行研究. 旺金晖等^[14] 研究了不同缺失比率、不同样本容量、不同参数先验分布下等级反应模型参数的 MCMC 估计.

可以看出, 缺失数据的存在确实会对项目参数产生各种影响. 然而, 已有研究在缺失数据的处理方面, 要么关注各因素对项目参数估计精度的影响, 要么关注的是随机缺失数据, 要么模型参数估计较复杂, 有的还需要复杂的自编程序, 不仅要预测变量与因变量之间的关系进行建模, 还要对缺失数据的生成机制进行建模. 对于一些数理基础较差的研究者来说, 使用这些复杂模型较为困难. 为此, 本文在已有研究的基础上, 拟通过简单的软件和操作解决一个简单但重要的问题, 即缺失数据比率和处理方式对非随机缺失数据能力参数估计精确性的影响到底如何, 为实践情境中如何相对准确地估计 MNAR 数据的能力参数提供借鉴.

1 缺失数据处理方法

缺失数据的处理方法有多种, 例如个案剔除法(listwise deletion)、成对删除法(pairwise deletion)、忽略不计(not presented, NP)、错误作答(incorrect, IN)、部分正确(fractionally correct, FR)、正确项目平均分替代法(corrected item mean substitution imputation, CM)、反应函数法(response function imputation, RF)、平均数插补法(mean imputation)、回归插补法(regression imputation)、多重插补法(MI)、EM 算法等^[4-5, 8]. 这些处理方法各有优缺点, 本研究拟选取 6 种易于理解的缺失数据处理方法: NP、IN、FR、CM、MI、EM. 在这 6 种方法中, NP、IN 和 FR 这 3 种方法不需要数据插补, 可以直接用 BILOG 软件进行参数估计. 数据插补涉及到基于观测数据的信息来推测缺失数据的可能取值. 例如, 可以根据被试在其他未缺失项目上的作答, 或其他被试在该缺失项目上的作答, 或是综合二者来推测被试在某一缺失项目上的可能取值.

NP 是把某一项目上的缺失数据当作此项目不包含在整个测验中, 就像没有对被试施测该项目一样. IN 是把缺失数据当作被试错误作答, 赋值为 0. 这也是大多数研究中所采取的缺失数据处理方法.

它假设被试尝试对项目进行作答, 但是却回答错误, 这样用错误作答来代替缺失数据假设被试不可能正确作答缺失项目, 而完全不考虑被试的能力. F. M. Lord^[15] 曾指出把缺失数据当作错误作答不是一种好方法, 虽然这是实践中经常使用的方法. 只有在缺失数据为 MCAR 时, IN 才不会产生有偏参数估计. FR 是把缺失数据当作部分正确作答, 例如, 在五选一选择题上的缺失数据, 将其赋分为 1/5. 在大规模国际学生评估项目 NAEP 中, 在处理选择题缺失作答时采用的 FR 方法^[6]. CM 是根据被试在其他项目的作答以及其他被试的作答来推算被试在缺失项目上的得分.

$$\begin{aligned} PM_i &= \sum_j x_{ij}/J_i, \\ IM_j &= \sum_i x_{ij}/I_j, \\ \tilde{x}_{ij} &= \left(\frac{PM_i}{\frac{1}{obs(i)} \sum_j IM_j} \right) IM_j, \end{aligned}$$

其中 x_{ij} 是被试 i 在项目 j 上的反应, 并且 x_{ij} 未缺失; J_i 为被试 i 未缺失的项目数; I_j 是在项目 j 上未缺失的被试数; \tilde{x}_{ij} 是被试 i 在项目 j 上的插补值; $obs(i)$ 是被试 i 未缺失的项目数. 对于 0、1 二项记分项目, 得到 \tilde{x}_{ij} 后, 根据贝努力分布 $N(1, \tilde{x}_{ij})$ 产生缺失数据的替换值.

MI 算法根据一定的模型对缺失数据进行 M 次插补, 从而形成 M 批完整反应矩阵, 分别进行参数估计, 最后把 M 次参数估计的结果进行整合. 根据相对效率的要求, MI 算法一般取 3 ~ 5 次插补^[16], 所以在本研究中取插补次数为 5. 在 IRT 模型中采用 MI 方法处理缺失数据有许多好处, 首先它可以从降低缺失数据不确定性带来的误差, 在保证统计检验力的前提下, 有较小的置信区间^[17]. EM 算法是一种良好的数据插补方法, 每次迭代包括 2 个步骤: E(求期望) 步, M(最大化) 步^[13, 18]. 张淑梅在研究 MAR 的缺失数据处理方式时, 得到用 EM 算法处理缺失数据获得的参数估计比 BILOG-MG 先插补再估计参数的方法具有更小的均方误和标准差^[13].

2 模拟设计

2.1 目的

本文拟采用 5×6 两因素组间设计, 考察缺失数据比率(5%、10%、20%、30%、50%) 和处理方法(IN、NP、FR、CM、MI、EM) 对非随机缺失数据能力参数估计准确性的影响, 以期在 IRT 背景下 MNAR 处理方式提供有用信息.

2.2 数据生成

首先,用 IRTLAB 模拟产生 50 批 500 个被试在 20 个 0-1 记分项目上作答数据,被试能力符合标准正态分布 $N(0, 1)$,采用 2 参数逻辑斯蒂克模型估计 20 个项目的项目参数和被试的能力参数,把完整作答数据下的估计参数作为真值.再次,产生不同比率的非随机缺失数据(5%, 10%, 20%, 30%, 50%),缺失数据产生过程参考 Holmes Finch 的研究.被试在某一项目上是否缺失与原反应有关,原反应作答正确的缺失概率较小,反之则大,以保证符合 MNAR 的缺失机制^[12].例如,在生成 5% 的缺失比率时:1) 根据完整反应矩阵生成同等大小的缺失概率矩阵,设置原得分为 1 的反应缺失概率为 2%,原得分为 0 的反应缺失概率为 8%;2) 生成同等大小的随机数矩阵,随机数矩阵符合均匀分布(0, 1),比较缺失概率矩阵和随机数矩阵,若缺失概率大于随机数,则完整反应矩阵相应位置的反应设为缺失,否则保持原反应不变,最终得到缺失比率为 5% 的缺失反应矩阵.

2.3 数据分析

采用 BILOG-MG 软件对模拟数据进行分析处理,其中 IN、NP、FR 这 3 种方法是 BILOG-MG 软件自带的缺失数据处理选项.其他 3 种方法经过插补形成完整反应数据:CM 方法根据计算公式采用 R 3.2.2 自编程序对缺失数据进行插补,EM 与 MI 方法采用 SPSS 20.0 对数据进行插补,形成完整作答数据.被试能力参数均采用 BILOG-MG 软件估计获得.

2.4 评价指标

采用 $BIAS_{abs}$ 、 $BIAS$ 、 $R_{(\theta, \hat{\theta})}$ 和 $RMSE$ 等 4 种指标揭示各种条件下被试能力参数的返真性,以反映缺失数据处理方法与缺失比率对 IRT 能力参数估计准确性的影响.

$$BIAS_{abs} = \sum_{m=1}^M \sum_{i=1}^I |\hat{\theta}_{im} - \theta_i| / MI,$$

$$BIAS = \sum_{m=1}^M \sum_{i=1}^I (\hat{\theta}_{im} - \theta_i) / MI,$$

$$RMSE = \sqrt{\sum_{m=1}^M \sum_{i=1}^I (\hat{\theta}_{im} - \theta_i)^2 / MI},$$

$$R_{(\theta, \hat{\theta})} = COV(\theta, \hat{\theta}) / \sigma_{\theta} \sigma_{\hat{\theta}},$$

其中 I 为被试数, M 为模拟重复次数,在本研究中为 50 次.

3 结果

3.1 不同处理方法与缺失比率下 $BIAS_{abs}$ 值的变化

$BIAS_{abs}$ 值反映被试能力真值与能力估计值之间差异的绝对值.表 1 是在不同缺失比率下不同缺失数据处理方法的 $BIAS_{abs}$ 值.由表 1 可知,从整体上看,缺失比率对被试能力参数估计准确性的影响趋势是一样的,不同的处理方法下,被试能力参数估计误差 $BIAS_{abs}$ 值均随着缺失比率的升高而增大.例如,在 5% 的缺失比率下,各方法的 $BIAS_{abs}$ 值在 0.044 0 ~ 0.103 6 范围内,当缺失比率到 50% 时,各方法的 $BIAS_{abs}$ 值上升到 0.298 8 ~ 0.436 1,并且缺失比率越高,不同方法 $BIAS_{abs}$ 值的差异也随之增大.值得注意的是,在本研究中,在各缺失比率下,误差值较大的缺失数据处理方法是 FR(除了 30% 和 50% 缺失比率下,CM 方法略高于 FR),这与以往研究所得结论不同(以往研究一般认为 IN 方法要明显比其他方法产生较大误差),而本研究中 IN 与 NP、MI、EM 方法相似,误差值都较小,并且在高缺失比率下,NP 方法的误差较小,CM 方法在较低缺失比率下误差小,但是随着缺失比率的升高,它的 $BIAS_{abs}$ 值反而超过其他几种方法.

表 1 在不同缺失比率下不同缺失数据处理方法的 $BIAS_{abs}$ 值

缺失比率/%	IN	NP	FR	CM	MI	EM
5	0.051 0	0.058 1	0.103 6	0.044 0	0.069 1	0.060 7
10	0.118 2	0.093 5	0.144 9	0.103 2	0.122 2	0.097 5
20	0.193 1	0.151 5	0.208 4	0.193 6	0.164 3	0.167 1
30	0.254 8	0.201 9	0.271 8	0.276 7	0.234 9	0.224 5
50	0.364 0	0.298 8	0.422 7	0.436 1	0.331 6	0.340 0

3.2 不同处理方法与缺失比率下 $BIAS$ 值的变化

不同处理方法与缺失比率下的 $BIAS$ 值如表 2 所示.总体而言,IN、NP、FR 这 3 种方法会过高估计被试能力参数,尤其是 IN 和 FR 方法,随着缺失比率的增大, $BIAS$ 值快速增高. CM、MI、EM 方法在不

同缺失比率下,时而高估被试的能力参数,时而又会出现低估的情况,例如,EM 方法在低缺失比率下(5%)会高估被试能力参数,在缺失比率大于 5% 时,则会低估被试的能力参数,并且随着缺失比率增大,偏差逐渐增大. NP 与 MI 方法在各缺失比率下

BIAS 值均小于 10^{-3} , 表现出较好的参数返真性.

表2 在不同缺失比率下不同缺失数据处理方法的 BIAS 值

缺失比率/%	IN	NP	FR	CM	MI	EM
5	0.372 7	0.177 4	0.547 1	0.311 9	0.123 2	0.117 6
10	1.204 7	0.207 8	0.884 7	-0.190 6	-0.017 5	-0.361 0
20	2.382 5	0.115 6	1.178 1	-0.603 3	0.045 8	-1.100 4
30	4.166 3	0.271 0	1.313 2	-1.251 7	0.189 8	-1.726 7
50	6.904 3	0.599 2	1.514 0	0.153 8	0.955 4	-2.742 4

注: 表中数值单位为 10^{-3} .

3.3 不同处理方法与缺失比率下 RMSE 值的变化

RMSE 一般表示变量的真值与估计值之间误差的变异性, 由表3可以看出, 与 BIAS 值相似, 随着缺失比率的升高, 各方法的 RMSE 值逐渐递增. 在各种方法的误差变异性方面, FR 的 RMSE 在大多数情况

下要高于其他方法, NP 的 RMSE 在大多数情况下是几种方法中较低的. CM 方法的 RMSE 在低缺失比率下与 IN、NP 方法相近, 但随着缺失比率的升高, 它的 RMSE 值快速递增, 在高缺失比率下甚至略高于 FR 方法. MI 与 EM 方法 RMSE 值变化较为平稳.

表3 在不同缺失比率下不同缺失数据处理方法的 RMSE 值

缺失比率/%	IN	NP	FR	CM	MI	EM
5	0.083 5	0.085 4	0.137 9	0.087 4	0.099 4	0.103 6
10	0.154 1	0.130 2	0.188 4	0.148 5	0.230 9	0.153 1
20	0.246 7	0.199 2	0.266 4	0.251 8	0.218 5	0.234 0
30	0.321 3	0.260 4	0.344 4	0.353 0	0.335 8	0.302 4
50	0.456 8	0.378 7	0.525 3	0.547 5	0.421 3	0.442 8

3.4 不同处理方法与缺失比率下 $R_{(\theta, \hat{\theta})}$ 值的变化

$R_{(\theta, \hat{\theta})}$ 是被试能力参数真值与估计值的皮尔逊相关系数, $R_{(\theta, \hat{\theta})}$ 越小表示能力参数估计值越接近其真值. 总体来看, 随着缺失比率增高, 各种方法的

$R_{(\theta, \hat{\theta})}$ 逐渐降低, 在缺失比率小于 30% 时, 各种方法的 $R_{(\theta, \hat{\theta})}$ 均在 0.9 以上. NP 方法在高缺失比率 (50%) 下, $R_{(\theta, \hat{\theta})}$ 仍保持在 0.911 5, 而其它方法则降为 0.90 以下, 如表4所示.

表4 不同处理方法与缺失比率下 $R_{(\theta, \hat{\theta})}$ 值的变化

缺失比率/%	IN	NP	FR	CM	MI	EM
5	0.995 9	0.995 7	0.988 8	0.995 5	0.994 2	0.993 7
10	0.985 9	0.990 0	0.979 1	0.987 1	0.968 3	0.986 4
20	0.963 5	0.976 3	0.958 7	0.963 0	0.971 5	0.968 5
30	0.937 3	0.959 2	0.932 6	0.926 9	0.931 8	0.947 8
50	0.868 6	0.911 5	0.845 1	0.818 0	0.889 7	0.889 4

3.5 不同缺失比率与处理方法下被试能力参数高估比率的变化

为了进一步探查被试能力水平不同时, 缺失数据比率与处理方法对能力参数估计准确性的影响机制. 根据被试的能力参数真值, 按照正态分布理论, 将被试能力分为高、中、低3组, 高分组被试能力真值大于等于1(约占总人数16%), 低分组被试能力真值小于等于-1(约占总人数16%), 其他被试属于中等能力组(约占总人数68%). 比率超过50%代表高估比率大, 低于50%代表低估比率大. 从各方法的总体趋势来看, 被试能力的高估比率均是高低能力组要大于中等能力组, 且被试能力高估比率

似乎与缺失比率无关, 如表5所示.

4 讨论

该文采用模拟研究探讨了数据缺失比率和处理方法对被试能力参数估计准确性的影响. 与已有研究不同的是, 该研究中的处理方法是一些操作较为简单的插补法或是删除法. 对于一些数理基础较差的研究者而言, 这些方法简单易行. 此外, 缺失数据的设计与实际较为符合, 原本作答错误的项目比原本作答正确的项目更容易缺失, 这同样与 MNAR 的定义相符.

4.1 缺失比率对 IRT 能力参数估计准确性的影响

在各种处理方法下,被试能力参数估计准确性随着缺失比率的增大而下降,并且下降趋势随着缺失比率增加幅度的增大而越来越明显,这与已有的研究结论是一致的^[7,19].这就要求研究者在数据收集阶段尽可能地保证数据的完整性,例如,通过合适的指导语增加被试对研究重要性的认识,增加他们的反应动机;或者增加调查回访程序,对被试的缺失数据进行补救.

表5 不同能力组被试在不同缺失数据处理方法和缺失比率下能力参数高估比率

处理方法	缺失比率/%	高	中	低
IN	5	0.734 4	0.171 0	0.791 8
	10	0.452 5	0.105 4	0.487 8
	20	0.357 6	0.083 3	0.385 6
	30	0.300 4	0.070 0	0.323 9
	50	0.230 5	0.053 7	0.248 5
NP	5	0.253 6	0.059 1	0.273 4
	10	0.322 5	0.075 1	0.347 7
	20	0.334 0	0.077 8	0.360 1
	30	0.318 6	0.074 2	0.343 4
	50	0.258 6	0.060 2	0.278 8
FR	5	0.442 3	0.103 0	0.476 8
	10	0.385 0	0.089 7	0.415 1
	20	0.271 0	0.063 1	0.292 2
	30	0.172 5	0.040 2	0.185 9
	50	0.040 3	0.009 4	0.043 5
CM	5	0.401 2	0.093 4	0.432 5
	10	0.649 6	0.151 3	0.700 3
	20	0.529 9	0.123 4	0.571 2
	30	0.499 8	0.116 4	0.538 8
	50	0.462 2	0.107 6	0.498 3
MI	5	0.264 3	0.061 5	0.284 9
	10	0.327 5	0.076 3	0.353 1
	20	0.353 7	0.082 4	0.381 3
	30	0.330 5	0.077 0	0.356 3
	50	0.264 6	0.061 6	0.285 2
EM	5	0.224 5	0.052 3	0.242 0
	10	0.349 2	0.081 3	0.376 4
	20	0.447 0	0.104 1	0.481 9
	30	0.479 8	0.111 7	0.517 3
	50	0.444 7	0.103 6	0.479 5

4.2 处理方法对 IRT 能力参数估计准确性的影响

IN 确实不是最佳的缺失数据处理方法,尽管这是大多数研究者经常采用,也是许多统计软件默认的缺失数据处理方法,这 F. M. Lord 的结论是一致的^[15].在估计项目参数时,IN 的处理方法产生较大的参数估计误差,但在能力参数估计时却会产生不

同的影响, $BIAS$ 值较大,其他几个指标与其他方法差异不大. IN 使得对被试的能力参数产生高估.出乎意料的是, FR 在处理非随机缺失数据时产生了较大的误差,这因为非随机缺失数据的产生与被试潜在特质有关,缺失作答等地赋予部分正确分数,与 MNAR 的假设相悖(与被试潜在特质等不可观测数据有关). NP 方法的 $BIAS_{abs}$ 、 $BIAS$ 值与 $RMSE$ 均较小, $R_{(\theta, \hat{\theta})}$ 较高,表现出能力参数估计较好的准确性和较小的变异性, NP 方法在估计被试能力参数时与那些基于模型的方法表现的一样好. CM 方法在低缺失比率下 $BIAS_{abs}$ 值较小,随着缺失比率的增加其 $BIAS_{abs}$ 值逐渐快速增大,在高缺失比率下其 $BIAS_{abs}$ 甚至超过 FR 方法,因为 CM 方法插补缺失值依赖于被试在其他项目上的作答和其他被试在缺失项目上的作答来插补缺失值,随着缺失比率的增加,这些有效值的比率降低,必然增大其误差. EM 算法的误差不是最大,但是和其他插补法相比似乎要稍微逊色,这是因为 EM 算法依赖于多元正态分布假设,而这似乎不适用于二级记分项目. MI 方法的 $BIAS_{abs}$ 值似乎更稳定,但是并不明显优于其他方法,这因为在 MI 法在插补数据之后需要取整,以适应参数估计软件,而本研究中记分方式为 0-1 记分,在取整时可能会使得插补值与取整之后相差较大^[12].总之,相比于复杂的基于模型的处理方法,简单的缺失数据处理方法也可以取得较好的能力参数估计准确性.

5 结论

1) 缺失比率会影响能力参数估计准确性,缺失比率越高,能力参数估计准确性越差,估计变异性越大.

2) NP 方法在处理非随机缺失数据时效果比其他非插补方法要好,在本文中,它在各个指标下均表现出良好的参数估计准确性.在 0-1 记分测试中不提倡采用 FR 方法处理非缺失数据.在低缺失比率下可以使用 CM 方法,而在高缺失比率下则不宜采用. MI 与 EM 方法表现出较好的参数估计稳定性,也是一种可以替换的非随机数据处理方法.

3) 未来研究可以关注混合题型时,数据缺失比率和处理方法对项目参数和能力参数的影响机制.

6 参考文献

- [1] 游晓锋,丁树良,刘红云. 缺失数据的估计方法及应用[J]. 江西师范大学学报:自然科学版, 2011, 35(3): 325-330.
- [2] Holman R, Glas C A W. Modeling non-ignorable missing-

- data mechanisms with item response theory models [J]. The British Psychological Society 2005, 58(1): 1-17.
- [3] Little R J A, Rubin D B. Statistical analysis with missing data [M]. Hoboken, NJ: John Wiley and Sons, Inc 2002.
- [4] Peugh J L, Enders C K. Missing data in educational research: a review of reporting practices and suggestion for improvement [J]. Review of Educational Research 2004, 74(4): 525-556.
- [5] Roth P L. Missing data: a conceptual review for applied psychologists [J]. Personal Psychology, 1994, 47(3): 537-560.
- [6] Rose N, von Davier M, Xu Xueli. Modeling nonignorable missing data with item response theory [EB/OL]. [2015-12-11]. <http://www.ets.org/Media/Research/pdf/RR-10-11.pdf>.
- [7] Newman D A. Missing data: five practical guidelines [J]. Organizational Research Methods 2014, 17(4): 372-411.
- [8] Kristin L, Sainani. Dealing with missing data and suggestion for improvement [J]. Review of Educational Research 2004, 74(4): 525-556.
- [9] 李斌, 李晓毅, 付志慧. IRT 框架下不可忽视缺失数据的 Bayes 估计 [J]. 沈阳师范大学学报: 自然科学版, 2015, 33(2): 216-220.
- [10] Pohl S, Gräfe L, Rose N. Dealing with omitted and not-reached items incompetence tests: evaluating approaches accounting for missing responses in item response theory models [J]. Educational and Psychological Measurement, 2014, 74(3): 423-452.
- [11] Köhler C, Pohl S, Carstensen C H. Taking the missing propensity into account when estimating competence scores: evaluation of item response theory models for nonignorable omissions [J]. Educational and Psychological Measurement 2015, 75(5): 850-874.
- [12] Finch H. Estimation of item theory parameters in the presence of missing data [J]. Journal of Educational Measurement 2008, 45(3): 225-245.
- [13] 张淑梅, 辛涛, 曾莉, 等. 2PL 模型的 EM 缺失数据处理方法研究 [J]. 应用概率统计 2011, 27(3): 241-255.
- [14] 汪金晖, 张淑梅, 辛涛. 缺失数据下等级反应模型参数 MCMC 估计 [J]. 北京师范大学学报: 自然科学版, 2011, 47(3): 229-234.
- [15] Lord F M. Maximum likelihood estimation of item response parameters when some responses are omitted [J]. Psychometrika, 1983, 48(3): 477-482.
- [16] Rubin, Donald B. Multiple imputation for nonresponse in surveys [M]. New York: John Wiley Press, 1987.
- [17] Kadengye D T, Ceulemans E, Noortgate W V D. Direct likelihood analysis and multiple imputation for missing item scores in multilevel cross-classification educational data [J]. Applied Psychological measurement, 2013, 31(8): 61-80.
- [18] Margot Peeters. How to handle missing data: a comparison of different approaches [J]. European Journal of Developmental Psychology 2015, 12(4): 377-394.
- [19] Schafer J L. Analysis of incomplete multivariate data [M]. Wiley-VCH: Godin Lytle Press, 1997.

The Effects of Missing not at Random Data to the Accuracy of Ability Parameter Estimation in IRT

KANG Chunhua, SUN Jinling, SUN Xiaojian, ZENG Pingfei *

(College of Teacher Education, Zhejiang Normal University, Jinhua, Zhejiang 321004, China)

Abstract: A simulation study is conducted to explore proper methods of handling missing not at random data in IRT context. First, generate response data of 500 subjects on 20 items, every item is scored by 0 or 1. Then, different percent of missing data were simulated, next 6 kinds of methods were used to deal with the missing data. Estimate the subjects' ability via the BILOG-MG software, comparing different methods with the following four criterias: *BIAS*, *BIAS_{abs}* and *RMSE*. It is shown that these methods exhibit varying degrees of effectiveness in dealing with MNAR. It is advisable for us to use the NP, MI, EM methods to handle with MNAR in IRT context.

Key words: the proportion of missing data; methods to deal with missing data; parameter estimation in IRT

(责任编辑: 冉小晓)