心理科学进展 2014, Vol. 22, No. 6, 1025–1035
Advances in Psychological Science DOI: 10.3724/SP.J.1042.2014.01025

·研究方法(Research Method)·

计划缺失设计——通过有意缺失让研究更高效*

王孟成 叶浩生

(广州大学心理与脑科学研究中心; 广州大学教育学院心理学系, 广州 510006)

摘 要 缺失值是社会科学研究中非常普遍的现象。全息极大似然估计和多重插补是目前处理缺失值最有效的方法。计划缺失设计利用特殊的实验设计有意产生缺失值, 再用现代的缺失值处理方法来完成统计分析, 获得无偏的统计结果。计划缺失设计可用于横断面调查减少(或增加)问卷长度和纵向调查减少测量次数, 也可用于提高测量有效性。常用的计划缺失设计有三式设计和两种方法测量。

关键词 缺失值; 计划缺失设计; 全息极大似然估计; 多重插补; 三式设计; 两种方法测量
分类号 B841

科学研究中, 缺失是个非常普遍的现象。毫不夸张地说, 凡是涉及到数据收集的情景就存在数据缺失现象。在社会科学研究中, 调查对象在完成问卷过程中出于各种原因(如, 疏忽大意、回避敏感问题)造成数据缺失。数据缺失对于研究者来说是个头痛的问题, 不仅损失了信息也增加了工作量和调查成本。传统的处理措施是直接将包含缺失的样本删去 (即列删法)。这种做法简便易行, 很多流行的统计分析软件(如 SPSS)默认此方法, 但是该方法对缺失值的假设条件要求较高, 不满足假设前提时会产生估计偏差。与此同时, 删去缺失样本也会消减样本总量进而影响统计功效。

随着方法的发展, 更高效的缺失值处理技术被提出, 其中全息极大似然估计法(Full Information Maximum Likelihood, FIML)和多重插补法(Multiple Imputation, MI)最为学者推崇(e.g., Graham, 2009, 2012; Schafer & Graham, 2002)。采用这两种方法进行缺失值分析, 在数据完全随机缺失(Missing Completely At Random, MCAR)和随机缺失(Missing At Random, MAR)条件下均能得到与完整数据一致的结果。

收稿日期: 2013-11-18
* 国家社会科学基金教育学青年项目: “心理健康教育的循证实践模式及本土化研究” (CBA130124)。
通讯作者: 叶浩生, E-mail: yehaosheng@yahoo.cn

计划缺失设计(planned missing design)是现代缺失值处理方法在研究设计上的新发展, 其通过独特的实验设计有意产生缺失值, 再用全息极大似然估计或多重插补来处理缺失值, 进而完成后续统计分析, 获得无偏的统计结果。计划缺失设计可用于横断面调查减少(或增加)问卷长度, 也可用于纵向调查减少测量次数。本文将就这一方法做全面系统的介绍, 具体安排如下: 首先介绍缺失值的机制, 接着介绍两种缺失值处理方法。第三、重点介绍有计划缺失设计用于横断面和纵向研究的常用形式。第四、简单介绍了如何估计缺失设计的统计功效。文章最后对缺失值和计划缺失设计做了简要评价。

1 缺失值的机制

缺失值的机制(Missing data mechanism)并非是造成缺失值的原因, 而是描述缺失值与观测变量间可能的关系(Schafer & Graham, 2002)。Rubin (1976)最早将缺失值的机制分为三类: 随机缺失, 完全随机缺失和非随机缺失(Missing At Non-Random, MANR)。

1.1 随机缺失(MAR)

当某变量出现缺失的可能性与模型中某些观测变量有关而与该变量自身无关时称作 MAR。换句话说, 造成缺失的可能性在于其他变量。例如, 在一次测试中, 如果 IQ 达不到最低要求的 100 分,

那么将不能参加随后的人格测验, 因此在人格测验上产生的缺失即为 MAR (与人格变量自身无关, 与 IQ 有关)。

目前, 缺少检验 MAR 的有效程序, 幸运的是严重违反 MAR 假设的情况相对较少(Schafer & Graham, 2002)。研究者推荐使用包含辅助变量(Auxiliary Variables, 与缺失变量相关的因素)的方法减少估计偏差并提高满足 MAR 假设的可能性(Collins et al., 2001; Schafer & Graham, 2002)。具体来说, 在分析缺失数据时将辅助变量纳入分析过程, 但辅助变量并不出现在模型中。

1.2 完全随机缺失(MCAR)

当某变量缺失值发生的可能性与该变量自身无关, 与其他变量也无关时称作 MCAR。换句话说, 某变量缺失值的出现完全是个随机事件, 因此可以将存在 MCAR 变量的数据看作是假定完整数据的一个随机样本(Rubin, 1976)。

数据是否是 MCAR 可以采用单变量 *t* 检验和 Little (1988)提出的多元 *t* 检验进行考察。其原理是, 如果变量 *X* 的缺失值是完全随机的, 那么在 *X* 上缺失和非缺失两组样本间在第二个变量 *Y* 上的均值差异是不显著的, 否则存在某种相关性。单变量和多元 *t* 检验可在 SPSS 上运行。然而均值差异比较并非能保证一定满足 MCAR, 因为在 MAR 和 MANR 条件下也能产生相等的均值(Enders, 2010)。

1.3 非随机缺失(MANR)

当某变量出现缺失值的可能性只与自身相关时称作 MANR。例如, 公司新录用了 20 名员工, 其中 6 名员工由于表现较差, 在试用期内被辞退, 年终表现评定中, 被辞退的 6 名员工的表现分缺失了, 此时的表现分数缺失即为 MANR (与其他变量无关, 与表现得分相关)。

不同的缺失值机制意味着不同的处理方法。MANR 与 MAR 和 MCAR 的情况不同, 因此下面讨论的方法主要集中在 MAR 和 MCAR 两种条件下, MANR 的处理方法可参见 Enders (2011)和 Muthén, Asparouhov, Hunter 和 Leuchter (2011)。

2 缺失值处理的现代方法

传统的方法存在种种不足, 新的方法也在不断发展, 其中最为研究者推崇的方法为多重插补(Multiple Imputation, MI)和极大似然估计(Enders, 2010; Graham, 2009, 2012; Schafer & Graham, 2002)。下面主要从非技术的角度介绍这两种方法, 统计取向的导引性文章见沐守宽和周伟(2011)。

2.1 全息极大似然估计

极大似然估计(Maximum Likelihood, ML)在处理缺失数据时称作全息极大似然估计, 意指使用所有观测变量的全部信息。FIML 同 ML 分析完整数据过程一样, 只是在计算单个对数似然值时使用全部完整信息而不考虑缺失值(Enders, 2010)。因此, ML 处理缺失值并非使用替代值将缺失值填补, 而是使用已知信息采用迭代的方式(通常采用 EM 算法)估计参数。FIML 在 MCAR 和 MAR 下产生无偏和有效的参数估计值。当在非正态分布时, FIML 可以使用参数校正统计量(如 S-Bχ^2, Bootstrap 等)。

FIML 法包含辅助变量的分析使用了 Graham (2003)提出的饱和相关模型(saturated correlates model), 即将辅助变量纳入模型, 同时允许辅助变量间、辅助变量与外生观测指标以及内生观测指标的测量误差相关[1]。某些潜变量建模软件如 Mplus 采用 FIML 分析带有辅助变量的模型时默认采用饱和相关模型(Muthén & Muthén, 1998–2012)。然而, 饱和相关模型纳入太多的辅助变量时, 会造成模型识别问题。通常选择与缺失变量关系密切的变量($r > 0.4$)作为辅助变量进入方程。

使用计划缺失设计的前提是掌握现代的缺失值处理技术, 上面只是从非技术的角度简单介绍了 FIML 处理技术, 下面以一个 17 个条目测量 3 个因子的验证性因素分析为例, 采用国际流行的结构方程建模软件 Mplus[2]来演示具体过程。

在 Mplus 中执行 FIML 分析与完整数据分析

---

[1]回归模型时, 只有辅助变量间、辅助变量与因变量、自变量间相关。

[2]目前绝大多数结构方程建模软件都支持 FIML, 但只有少数软件支持 MI 功能。为了分析工具的统一, 本文选择两种功能均具备的 Mplus 作为演示工具。Mplus 还可以通过蒙特卡洛模拟来估计计划缺失设计统计功效损失情况(Enders, 2010)。另外, Mplus 在国内外潜变量建模领域的应用越来越流行(e.g., 毕向阳, 马缨, 2012; 王孟成, 2014; Wang, Elhai, Dai., & Yao, 2012; 叶宝娟, 温忠麟, 2012)。

表 1　FIML 处理缺失值的 Mplus 语句

```
TITLE: CFA Model With Missing Data;
DATA:FILE IS data with Missing.dat; !设置数据文件
VARIABLE: NAME are gender age y1-y17; !定义数据文件里的变量
          USEVARIABLES= y1-y17; !选择使用的变量
          MISSING=ALL (9); !定义缺失值标签
         AUXILIARY = (m) gender age; !设置自动包含辅助变量, (m)指定缺失分析的辅助变量, 如果是类别变
          量在变量名后加“(c)”。
ANALYSIS:ESTIMATOR = ML; !非正态时可以选用稳健极大似然估计 MLM;
          INFORMATION = Observed; !Mplus 提供 Expected 和 Observed 两种信息矩阵, 两者的区别在于计算
          标准误时依据的缺失值机制不同, 期望矩阵的要求是 MCAR, 观测矩阵依据的是 MAR, 所以通常
          使用的 Observed 矩阵(Enders, 2010)。
MODEL:  F1 by y1-y5;!定义测量模型
        F2 by y6-y12;
        F3 by y13-y17;
OUTPUT: STANDARDIZED; !要求报告标准化结果
```

注：!后为语句说明, 运行时软件会忽略这些信息。

过程基本相同, 如果数据文件中包含缺失值需要在输入语句中标注, 一旦标明 Mplus 自动执行 FIML。具体来说, 在“VARIABLE”命令下加入“MISSING=ALL (9)”即数据文件中所有变量的缺失值都使用数值 9 来表示。带有详细解释的完整命令呈现在表 1 中。

2.2　多重插补法

多重插补法最早由 Rubin (1987)提出, 假设在数据随机缺失情况下, 用两个或更多能反映数据本身概率分布的值来填补缺失值。一个完整的 MI 包含三步：数据插补(imputation phase), 计算(analysis phase)和汇总(pooling phase)。数据插补是最关键的一步, 对每一个缺失数据插补 m (m > 1)次。每次插补将产生一个完整数据集, 共产生 m 个完整数据集。第二步, 对每一个完整数据集采用标准的数据分析方法进行分析。第三步, 将每次分析得到的结果进行综合, 得到最终的统计推断。

根据数据缺失机制、模式以及变量类型, 可分别采用回归、预测均数匹配(Predictive Mean Matching, PMM)、倾向得分(Propensity Score, PS)以及常用的马尔可夫链蒙特卡罗(Markov Chain Monte Carlo, MCMC)等方法进行填补。MI 要求数据缺失为 MAR, 如果采用 ML 估计同样要求数据分布符合多元正态分布假设, 但研究发现违反多元正态假设对 MI 参数精确性影响不大(Demirtas, Freels, & Yucel, 2008)。另外一个影响估计精确性的因素是缺失率, Demirtas 等(2008)的研究发现缺失率高达 25%仍能得到精确的参数估计结果。

在具体使用 MI 时需要考虑 m 的次数。理论上, m 的值越大估计越精确, 但太大的数量会增加计算负荷。模拟研究发现, 当 m 取 20 时在多数情况下是合适的(Graham, Olchowski, & Gilreath, 2007)。在实践中, m 通常取 200~1000 之间。另外, 同 FIML 一样, MI 插补时(仅在第一步)也可以纳入辅助变量用于提高参数估计的精确性。但与 FIML 不同, MI 纳入辅助变量的数量不受限制。

在 Mplus 中执行 MI 需要两步[3], 第一步数据插补, 第二步使用第一步插补的数据集估算并输出汇总结果, 两步的 Mplus 语句分别呈现在表 2 中。

与 FIML 相比 MI 的语句要复杂不少, 多了数据插补部分。其他 MI 软件的分析过程基本类似, 先执行数据插补再对插补数据做分析汇总。

2.3　两种方法的比较与应用选择

两种方法都是目前解决缺失值问题最有效的方法(e.g., Graham, 2009; Schafer & Graham, 2002),

[3] 在最新的 Mplus7 中(手册第 11 章), MI 只需一步即可实现。这里为了方便理解仍做两步演示。

表 2 多重插补法处理缺失值的 **Mplus** 语句

```
    TITLE: This is an example of MI for first step;
DATA: FILE IS data with Missing.dat;
VARIABLE: NAME are gender age y1-y17;
          USEVARIABLES= y1-y17;
          MISSING=ALL (9);
          AUXILIARY = (m) gender age;
DATA IMPUTATION:
          IMPUTE = y1-y17; !设置存在缺失值的变量进行插补
          NDATASETS = 500; !设置 m=500, 软件默认值为 5;
          SAVE = dataimp*.dat; !设置插补的 50 个数据集保存文件名。
ANALYSIS: TYPE = basic;
          bseed = 79566; !数据扩张(data augmentation)时的种子
          chains = 1; !数据扩张时的贝叶斯链
```

```
TITLE: This is an example of MI for second step
DATA: FILE IS dataimplist.dat; !使用第一步保存的插补数据文件集;
TYPE = imputation; !说明数据类型为插补数据
VARIABLE: NAME are gender age y1-y17;
          USEVARIABLES= y1-y17;
ANALYSIS: ESTIMATOR = ML;
          INFORMATION = observed; !标准误基于观测信息矩阵;
MODEL:    F1 by y1-y5;
          F2 by y6-y12;
          F3 by y13-y17;
OUTPUT:   STANDARDIZED; !输出标准化结果
```

但在实际应用中如何选择是广大应用研究者最关心的问题。理论上, 当 MI 采用先验分布同时插补的次数无限大时, 两种方法得到的结果是等价的(Enders, 2010)。在实际应用中, 两种方法所得结果也非常接近。由于 MI 法的精确性取决于插补的次数, 在实践中插补的次数不可能无限大, 次数太大会消耗更多的分析时间, 而 FIML 则不存在这一问题。另外, 多数流行的结构方程建模软件(少数软件除外, 如 Mplus)并不支持 MI, 需要专门的软件实现(如 NORM; Schafer, 1997)。即使包含 MI 功能的 Mplus 软件, 在分析时也要分 2 步完成(在最新的版本中只需一步), 而 FIML 是多数结构方程软件默认功能。综合来看, 在实践中选择 FIML 比 MI 更简便。

3 计划缺失设计

问卷调查是心理学研究的主要形式和手段之一。多数研究需要使用多个问卷或量表进行调查, 而太长的问卷往往造成参与者认知疲劳甚至诱发抵触情绪, 影响问卷回答的质量, 威胁研究效度。如果能够通过某种研究设计主动产生 MAR 和 MCAR 缺失数据, 再用现代缺失值处理方法来获得无偏估计, 这些问题便迎刃而解了。计划缺失设计就是这样一种方法, 通过特定的研究设计让参与者完成部分问卷(产生 MCAR 数据)即可获得完整问卷所能提供的信息, 主流的设计方法有三式设计(3-form design)及其扩展的多式设计(multi-form design)和两种方法测量(two-method measurement)。

3.1 三式及多式设计

3.1.1 三式及多式设计

三式设计(Graham, Hofer, & MacKinnon, 1996; Graham, Taylor, Olchowski, & Cumsille, 2006)是计划缺失设计方法中最简单易行也是最通用的方

法。三式法要求在完整版问卷的基础上建立 3 种形式的问卷, 每种形式之间有 2/3 的条目是重叠的(见表 3)。在调查实施过程中只要求每个参与者回答一种形式问卷(类似完全随机试验设计), 因此可以增加整体问卷长度或缩短回答时间。

表 **3**　三式设计模型

形式	问卷或条目集			
	X	A	B	C
1	O	M	O	O
2	O	O	M	O
3	O	O	O	M

注：O 表示观测; M 表示缺失。

表 3 呈现了三式设计法的问卷组合形式。假如某项调查需要使用 4 个量表或问卷分别为 X, A–C。其中问卷 X 为研究的重点, 因此在三种形式问卷中均包含, 而其他 3 个问卷在 3 种形式问卷中只包含其中 2 个, 即 A&B, A&C 和 B&C。如果有 600 人接受调查, 每种形式的问卷均有 200 人参与。

如果调查所要使用的问卷较多, 可以将其划分为更多的组合形式, 常用的有六式(表 4)和十式。

表 **4**　六式设计模型

形式	问卷或条目集					
	X	A	B	C	D	E
1	O	O	O	M	M	O
2	O	O	M	O	M	O
3	O	M	O	O	M	O
4	O	O	M	M	O	O
5	O	M	O	M	O	O
6	O	M	M	O	O	O

注：O 表示观测; M 表示缺失。

上述三式或六式设计是在问卷或量表水平上进行, 如果在条目水平上确定问卷组合形式其思路与之类似, 只是在挑选问卷条目组成 X 时需要考虑条目质量和内容效度(Moore, 2012)。由于 X 部分是研究的核心, 所以在确定时需要充分考虑内容代表性和条目质量。X 的产生和结构方程模型中的打包(item parceling)技术非常类似, 具体的措施可参考项目打包的策略(王孟成, 2014; 吴艳, 温忠麟, 2011)。

下面以一个例子来说明选择条目组合的设计形式。假如存在 2 个量表各包含 10 个条目, 在此基础上建立 3 种形式问卷, 具体过程如下：首先, 根据理论或先前研究从每个量表中选择 4 个条目组成 X (表 5)。由于缺失值估计需要根据已知信息推算缺失信息, 所以不同形式问卷之间的重叠内容越多[4], 缺失值估计的精确性越高, 这里从每个问卷里选择 4 个因子负荷(根据以往研究或预实验结果)较高的条目组成 X。第二, 将余下的 6 个条目分成 3 个组(A, B, C), 每组 2 个条目。第三, 将 X 和 3 个组(A, B, C)组合成 3 个略减版问卷, 即 X&A&B, X&A&C 和 X&B&C (表 6)。

通过上面的介绍不难发现, 计划缺失设计类似于传统的实验设计, 不同的问卷形式类似实验条件, 不同的参与者随机确定完成某种形式的问卷, 所以计划缺失设计产生的缺失数据为完全随机缺失, 采用现代缺失值处理方法可以得到无偏的参数估计结果。

与传统设计相比, 三式设计具有如下优点：首先, 参与者回答的问卷条目更少了, 减轻了参与者的负担, 避免了认知疲劳, 提高了数据质量, 同时也减少由疲劳造成的非计划缺失(Raghunathan & Grizzle, 1995)。第二, 每种形式的问卷省去 1/3 的条目, 减少了印刷和数据录入成本(Graham et al., 2006)。第三, 可以增加总问卷长度获取更多信息。由于时间限制, 原完整版问卷最多只能包含 200 个条目, 现在使用三式设计可以将数量提升至 250 个, 获得的信息量增加了。

计划缺失设计产生的是完全随机缺失数据, 使用现代缺失值处理所得参数估计结果是无偏的, 其唯一的不足是可能损失统计功效(Graham et al., 2006; Rhemtulla & Little, 2012)。然而功效的损失与缺失率关系不大, 主要取决于变量间预期相关效果量的大小。如果研究假设的相关不强时, 在产生 3 种形式问卷时, 各形式问卷之间重叠比例尽量大, 同时将预期相关较小的变量应同时放到 X 中。

[4] 如果研究中, 与其他变量预期相关较低的变量应该放到 X 项目组合里, 如果放到 A-C 里允许缺失则会损失更多统计功效(e.g., Rhemtulla & Little, 2012)。

表 5 条目水平的三式设计：条目选择

	条目集					条目集			
量表 1	X	A	B	C	量表 2	X	A	B	C
a1	X				b1		A		
a2	X				b2		A		
a3	X				b3	X			
a4	X				b4	X			
a5		A			b5	X			
a6		A			b6	X			
a7			B		b7			B	
a8			B		b8			B	
a9				C	b9				C
a10				C	b10				C

表 6 条目水平的三式设计：条目组合

形式		条目集
问卷 1	X&*A*&**B**	a1-a4 b3- b6 *a5 a6 b1 b2* **a7 a8 b7 b8**
问卷 2	X&*A*&C	a1-a4 b3-b6 *a5 a6 b1 b2* a9 a10 b9 b10
问卷 3	X&**B**&C	a1-a4 b3-b6 **a7 a8 b7 b8** a9 a10 b9 b10

3.1.2 三式和多式设计用于追踪研究

三式设计可以方便的推广到纵向追踪研究(Graham et al., 2006; Graham, 2012; Mistler & Enders, 2012; Rhemtulla & Little, 2012)。纵向追踪研究的特点是采用同一测量工具对同一样本间隔一段时间反复测量多次。追踪研究中，由于每次问卷相同，反复测量很容易让参与者产生厌倦和练习效应。采用计划缺失设计可以让参与者少参与一次或数次，减少测量误差，提高数据质量。

用于追踪研究时，三式设计将总样本分成 3 组(或 6 组)，在每次追踪调查时选择某些组不参与该次调查。下面以六式设计用于有 6 次测量的追踪研究为例(见表 7)说明三式设计用于追踪研究的特点。首先将全样本随机分成人数均等的 6 个组，然后按照表 7 的示意确定每组是否参与每个时点的调查。

总的来说该设计具有如下特点：(1)高效。全部样本只参加了 4 次的追踪调查，中间 4 次调查只有一半的参与者参与，但最后获得了 6 次测量的数据。(2)省时。较少的重复测量获得较长的追踪时段。假设每次重测间隔半年，每组参与者只需参加 4 次就可以观测 2.5 年的发展轨迹。(3)经济。假如每组有 100 名参与者，每次每人的调查费用为 20 元(问卷打印，录入，小礼品)，采用传统的完整样本研究设计，全部过程要花费 7.2 万元(100 人×6 次×6 组×20 元)，而采用计划缺失设计只需要 4.8 万元(100 人×4 次×6 组×20 元)，经费节省了 1/3。(4)有效。采用表 7 的设计，尽管数据点减少了 33%，但统计功效(statistical power, 1-β)并不会削减。如果以获取完整数据(如 250 人 6 次测量共 1500 个数据点)的代价来计算，采用计划缺失数据可以对更多参与者进行施测(375 人 4 次测量共 1500 个数据点)，统计功效不但不会消减，反而会增加统计(样本量扩大了)。

表 7 六式纵向设计

样本	测量时点					
	T1	T2	T3	T4	T5	T6
1	O	O	O	M	M	O
2	O	O	M	O	M	O
3	O	M	O	O	M	O
4	O	O	M	M	O	O
5	O	M	O	M	O	O
6	O	M	M	O	O	O

3.1.3 用于追踪研究的另外一种形式

上面介绍的三式设计用于追踪研究时采用的问卷还是完整的问卷，只是在重复调查时安排某些组不参加该次调查。如果将不同形式的问卷(略减版问卷)用于追踪调查有 2 种常用的设计，见表 8。

表 **8**　纵向三式设计模式

样本	T1	T2	T3	T1	T2	T3
1	问卷 1	问卷 1	问卷 1	问卷 1	问卷 2	问卷 3
2	问卷 2	问卷 2	问卷 2	问卷 2	问卷 3	问卷 1
3	问卷 3	问卷 3	问卷 3	问卷 3	问卷 1	问卷 2

如果研究的目的是为了揭示单个变量跨时间的相关, 各组每次测量使用相同的略减版问卷即可, 如表 8 的左半部分。这种策略的目的是让每次测量间的变量最大程度的重叠达到相关最大化。如果研究的目的是揭示不同变量间的关系, 同时重测效应最小化, 那么每个测量时点应使用不同的问卷, 即采用表 8 右半部分设计。

3.1.4　交叉序列设计(Cross-Sequential Design)

交叉序列设计是纵向研究最常用的设计方法之一(Duncan, Duncan, & Hops, 1996), 该方法允许研究者追踪少数几个年龄队列来检验变量随年龄变化的发展轨迹。表 9 陈列了交叉序列设计产生缺失值的类型, 通过对 4 个年龄(12~15 岁)队列连续追踪 3 年可以评估 6 个年龄段(12~17 岁)的发展状况。表 9 的数据可以采用潜增长曲线模型(Latent Growth Curve Model, LGCM)分析, 但不能用于相关分析。比如, 不能计算 12 岁和 17 岁之间的相关系数, 原因是两列变量均缺失。

表 **9**　交叉序列设计

队列	测量点的年龄					
	12	13	14	15	16	17
12	O	O	O	M	M	M
13	M	O	O	O	M	M
14	M	M	O	O	O	M
15	M	M	M	O	O	O

3.2　两种方法测量设计(Two-Method Measurement Design)

两种方法测量设计适用于某个建构(construct)的测量存在两种以上的方法或工具。通常, 存在这样一种情况, 有些方法简便易行、成本低廉但存在较大的测量偏差; 而另一些方法, 复杂精细、成本昂贵但误差小。在传统的设计中, 研究者权衡利弊择其一。在预算一定的情况下, 要么采用低廉的方法获取大量低质样本达到统计功效最大化, 要么采用昂贵的方法获得有限的高质样本。例如, 某研究想揭示学前儿童攻击行为与同伴关系之间的关系时, 研究可以采用教师评定和行为观察两种方法。前者测量方便快捷、可以获得大量样本, 但精确性不足(教师不能留意到所有儿童以及准确记住他们的行为); 而行为观察法较为精确, 但耗时、成本高, 可获得的样本量也有限。

如果能取两种方法之长, 即获得大量低质样本又获得少量高质样本则两全其美, 两种方法测量就是这样一种两者兼顾的方法。采用低廉不精确的方法获取大样本, 同时在该样本中选取部分样本采用昂贵精确的方法(未接受的样本作为缺失值处理)。例如, 采用教师评定法测量 1000 名学生, 然后在这 1000 名中抽取 200 名接受行为观察。这样就可以得到 200 个完整的样本和 800 个缺失行为观察数据的样本。由于低廉的方法存在测量偏差, 因此需要将偏差提取出来, 通常是在结构方程模型框架内采用共同方法偏差的矫正程序来处理(Graham et al., 2006; Graham, 2012; 王孟成, 2014), 如图 1 所示的偏差校正模型。

采用两种方法测量需要满足几个前提：(1)低廉的方法存在系统偏差, 昂贵的方法是无偏的。换句话说, 昂贵的方法比廉价的方法更精确。如果存在低廉而又无偏的方法, 一切问题都简单了。(2)两种方法或工具测量的是同一构念。(3)研究的焦点在于群体水平。(4)样本量最够大。至少要有 125 个样本接受 2 种方法的测量(Rhemtulla & Little, 2012)。

4　缺失设计统计功效估计

在具体研究中, 由于研究条件各不相同, 采用的缺失设计也不尽相同, 所以在设计实施前需要考虑所采用的缺失设计的统计功效。在多变量分析中, 统计功效的计算非常复杂, 无法采用常规的统计功效进行估计。常用的方法有假设检验法(Satorra & Sari, 1985)和蒙特卡洛模拟法(Muthén & Muthén, 2002)。假设检验法计算过程繁琐, 在一般结构方程软件中不能直接计算, 而

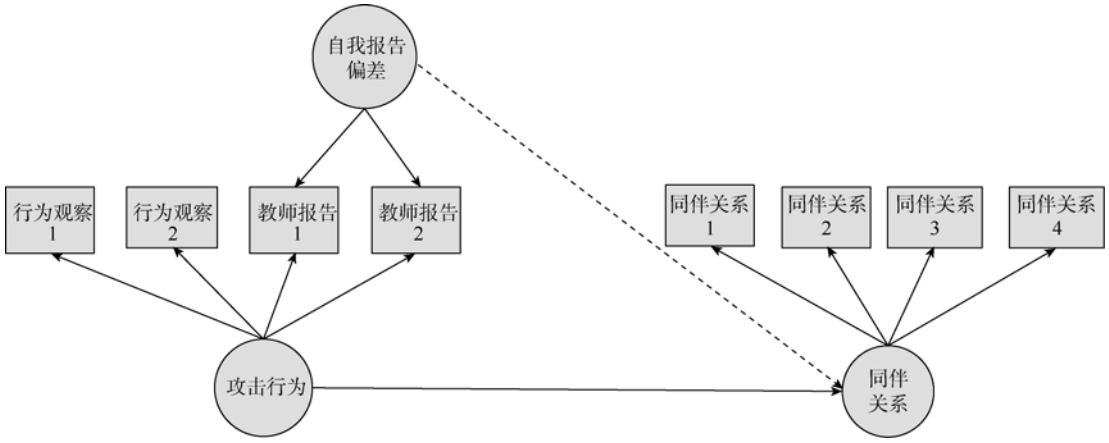

图 1 偏差校正模型(包含虚线的模型为完全偏差校正模型)

蒙特卡洛模拟法[5]在通常的 SEM 软件中可以方便的实现。下面以 2 因子相关验证性因子分析模型为例说明采用 Mplus 软件执行蒙特卡洛模拟法统计功效分析的过程。

在 Mplus 软件中采用蒙特卡洛模拟法估计统计功效非常方便。在 MONTECARLO 命令下定义模拟研究的条件, 通常涉及样本量、变量数、重复取样数等。缺失值模拟时需要使用 PATMISS 和 PATPROBS 分别设定缺失类型和每种缺失类型的缺失率。接着在 MODEL POPULATION 下面设定模拟所依据的总体样本参数。在 ANALYSIS 中确定分析所采用的方法和拟合标准等。具体见表 10, 更多设置见(Muthén & Muthén, 1998-2012)

表 10 中示例的模型为一个 2 因子相关验证性因子分析模型, 每个因子有 3 个测量指标, 因子负荷设定为 0.7, 因子相关设为 0.3。假设采用完整数据的设计(无缺失值设计)时的样本量为 250 人(250×6=1500), 而采用表 4 中的缺失设计时可以收集 375 人(375×4=1500)。以检验因子相关系数 0.3 的统计功效为例, 采用完整数据设计时的统计功效为 0.957(1000 次重复抽取的样本中有 957 个样本可以正确探测因子相关显著不为 0), 而采用缺失设计时的统计功效为 0.955, 与完整数据时的功效相差无几。其他类型的模型思路基本一致, 在实际研究中如何设定总体参数值是件困难的事情, 需要研究者结合过往研究结果和预调查结果。

5 总结与评价

缺失值是科学研究中非常普遍的现象。由于传统的处理方法会产生估计偏差而饱受诟病, 新的有效的方法在 MAR 和 MCAR 下产生无偏估计而受推崇(Graham, 2009, 2012; Schafer & Graham, 2002)。有效的处理方法被提出多年, 并可通过软件方便实现(e.g. SPSS, SAS, Mplus), 但传统的方法仍然普遍使用(Jeličić, Phelps, & Lerner, 2009 ; Peugh & Enders, 2004)。原因在于两个方面：一方面传统的方法非常方便, 很多统计软件默认或提供这些功能。另一方面, 研究者没有掌握新的处理方法, 甚至对当前有效的方法并不了解、知晓。

国际心理学期刊对缺失值处理的要求越来越严格。笔者投往国外心理学期刊的论文在审稿过程中多次遇到缺失值处理的审稿意见, 要求说明缺失值是否存在以及处理的方法。目前, 国内心理学期刊尚未对缺失值处理提出明确要求, 但随着国内心理学与国际接轨的步伐日益加快(如国内第一本英文心理学期刊 PsyCh Journal 的出版), 不远的将来, 缺失值处理问题将会在国内心理学者间达成共识。

随着新的方法日益普及, 缺失值问题已不再是研究者的噩梦, 而计划缺失设计则通过特殊的实验设计产生完全随机的缺失数据, 再用现代的

[5] 关于结构方程建模领域如何进行蒙特卡洛模拟研究, 可参见 Paxton et al.(2001)和温忠麟、刘红云、侯杰泰(2012)。

缺失值处理方法来完成统计分析，获得无偏的统计结果。本文介绍了当前两种常用的计划缺失设计，三式及其扩展的多式设计和两种方法测量。三式设计可用于横断面调查减少问卷长度，也可用于纵向调查减少测量次数。两种方法测量兼具高效昂贵的方法和低廉低质方法的优点，通过结

表 10　缺失设计统计功效蒙特卡洛模拟估计[6]

```
    TITLE: This is an example of power analysis using Monte Carlo simulation
MONTECARLO:!设置 Monte Carlo 模拟的参数
  names are y1-y6; !定义变量
  nobservations = 375; !设置样本量
  nreps = 1000; !设置重复抽取样本数
  seed = 56798; !设置蒙特卡洛模拟抽样的起始种子
  patmiss = !设置缺失的类型，括号内的 1 表示缺失, 0 表示未缺失
  y1(0) y2(0) y3(0) y4(1) y5(1) y6(0)|
  y1(0) y2(0) y3(1) y4(0) y5(1) y6(0)|
  y1(0) y2(1) y3(0) y4(0) y5(1) y6(0)|
  y1(0) y2(0) y3(1) y4(1) y5(0) y6(0)|
  y1(0) y2(1) y3(0) y4(1) y5(0) y6(0)|
  y1(0) y2(1) y3(1) y4(0) y5(0) y6(0);
  patprobs = !设置每种缺失类型的缺失比例
  .166666|.166666|.166666|.166666|.166666|.166666;
MODEL POPULATION: !设置模拟所依据的总体参数
  f1 by y1*.7 y2*.7 y3*.7;
  f2 by y4*.7 y5*.7 y6*.7;
  f1@1 f2@1;
  f1 with f2*.3;
  y1*.51 y2*.51 y3*.51;
  y4*.51 y5*.51 y6*.51;
ANALYSIS:
  coverage =.0000001; !设置收敛标准
    MODEL: !设置分析的模型
  f1 by y1*.7 y2*.7 y3*.7;
  f2 by y4*.7 y5*.7 y6*.7;
  f1@1 f2@1;
  f1 with f2*.3;
  y1*.51 y2*.51 y3*.51;
  y4*.51 y5*.51 y6*.51;
```

[6] 表中软件输出结果可与作者联系获得。

构方程模型减少方法偏差，提高研究结论的效度。希望本文能为国内学者了解、熟悉新的缺失值处理方法以及计划缺失设计提供一点有用的信息。

致谢：感谢芬兰 University of Turku 的 An Yang 博士和中国政法大学社会学院毕向阳博士对本文初稿提出的建议。同时感谢本文匿名审稿人对本文提出的建设性意见。

参考文献

毕向阳, 马缨. (2012). 重大自然灾害后社区情境对心理健康的调节效应—— 基于汶川地震过渡期两种安置模式的比较分析. *中国社会科学, 32*(6), 151–169.

沐守宽, 周伟. (2011). 缺失数据处理的期望-极大化算法与马尔可夫蒙特卡洛方法. *心理科学进展, 19*, 1083–1090.

王孟成. (2014). *潜变量建模与Mplus应用: 基础篇*. 重庆: 重庆大学出版社.

温忠麟, 刘红云, 侯杰泰. (2012). *调节效应和中介效应分析*. 北京: 教育科学出版社.

吴艳, 温忠麟. (2011). 结构方程建模中的题目打包策略. *心理科学进展, 19*, 1859–1867.

叶宝娟, 温忠麟. (2012). 测验同质性系数及其区间估计. *心理学报, 44*, 1687−1694.

Collins, L. M., Schafer, J. L., & Kam, C.-M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods, 6*, 330–351.

Demirtas, H., Freels, S. A., & Yucel, R. M. (2008). Plausibility of multivariate normality assumption when multiply imputing non-Gaussian continuous outcomes: A simulation assessment. *Journal of Statistical Computation and Simulation, 78*, 69–84.

Duncan, S. C., Duncan, T. E., & Hops, H. (1996). Analysis of longitudinal data within accelerated longitudinal designs. *Psychological Methods, 1*, 236–248.

Enders, C. K. (2010). *Applied missing data analysis*. New York: The Guilford Press.

Enders, C. K. (2011). Missing not at random models for latent growth curve analyses. *Psychological Methods, 16*, 1–16.

Graham, J. W. (2003). Adding missing-data relevant variables to FIML-based structural equation models. *Structural Equation Modeling, 10*, 80–100.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60*, 549–576.

Graham, J. W. (2012). *Missing data analysis and design*. New York: Springer.

Graham, J. W., Hofer, S. M., & MacKinnon, D. P. (1996). Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures. *Multivariate Behavioral Research, 31*, 197–218.

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science, 8*, 206–213.

Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods, 11*, 323–343.

Jeličić, H., Phelps, E., & Lerner, R. M. (2009). Use of missing data methods in longitudinal studies: The persistence of bad practices in developmental psychology. *Developmental Psychology, 45*, 1195–1199.

Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*(404), 1198–1202.

Mistler S. A., & Enders, C. K. (2012). Planned missing data designs for developmental research. In B. Laursen, T. D. Little, & N. A. Card (Eds.), *Handbook of developmental research methods* (pp. 742–754). New York: Guilford Press.

Moore, E. W. G. (2012). *Planned missingness study design: Two methods to developing the study survey versions*. Retrieved November 3, 2013, from http://crmda.ku.edu/resources/kuantguides/23.Planned_Missingness_Survey.pdf

Muthén, B., Asparouhov, T., Hunter, A. M., & Leuchter, A. F. (2011). Growth modeling with nonignorable dropout: Alternative analyses of the STAR*D antidepressant trial. *Psychological Methods, 16*, 17–33.

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (7th ed.). Los Angeles: Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling, 9*, 599–620.

Paxton, P., Curran, P. J., Bollen, K. A., Kirby, J., & Chen, F. N. (2001). Monte Carlo experiments: Design and implementation. *Structural Equation Modeling, 8*, 287–312.

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research, 74*, 525–556.

Raghunathan, T. E., & Grizzle, J. E. (1995). A split questionnaire survey design. *Journal of the American Statistical Association, 90*, 54–63.

Rhemtulla, M., & Little, T. D. (2012). Planned missing data designs for research in cognitive development. *Journal of*

Cognition and Development, 13, 425–438.

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

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Hoboken, NJ: Wiley.

Satorra, A., & Saris, W. E. (1985). Power of the likelihood ratio test in covariance structure analysis. *Psychometrika, 50*, 83–90.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. Boca Raton, FL: Chapman & Hall.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147–177.

Wang, M. C., Elhai, J. D., Dai, X. Y., & Yao, S. Q. (2012). Longitudinal invariance of posttraumatic stress disorder symptoms in adolescent earthquake survivors. *Journal of Anxiety Disorders, 26*, 263–270.

Planned Missing Data Design: Through Intended Missing Data Make Research More Effective

WANG Mengcheng; YE Haosheng

(*Center for Psychology and Brain Science*; *Department of Psychology of Education College*, *Guangzhou University*, *Guangzhou* 510006, *China*)

Abstract: Missing data is a very common phenomenon in social science research. Among most existing statistical analysis about missing data, Full-information maximum likelihood estimation and multiple imputation are recommended as most common approach at present for handling missing data. Planned missing data designs use special research design which create completely random missing data, and then employ modern missing data estimation techniques to get unbiased parameters and maximize statistical power in the process. Planned missing data designs can be used in cross-section survey to increase questionnaire items or reduce respondent burden. In longitudinal survey to shorten measure occasion, moreover, enrich validities. There are two types of planned missing data designs: 3-form design and two-method measurement design.

Key words: missing data; planned missing data designs; full-information maximum likelihood estimation; multiple imputation; 3-form design; two methods measurement design