

增长混合模型： 分析不同类别个体发展趋势*

王孟成 毕向阳 叶浩生

提要:追踪研究可以展现个体随时间的变化趋势、个体间差异并更好地揭示变量之间的因果关系,日益受到社会科学研究者的青睐,处理追踪数据的统计模型也日新月异。传统的分析模型假设样本同质,然而当此假设不满足时,就需要考虑群体的异质性问题。目前,处理异质性样本增长问题的模型主要有潜类增长模型(或组基增长模型)和增长混合模型。增长混合模型构成一般的形式,在一定条件下传统的潜在增长曲线模型和潜类增长模型均是其特例。沿着增长混合模型的发展路径,本文从应用性角度介绍了上述三个模型各自的特点、优势和 Mplus 软件实现,并通过一个实例演示了分析过程。最后,对当前研究存在的问题和将来的研究方向做了简要讨论。

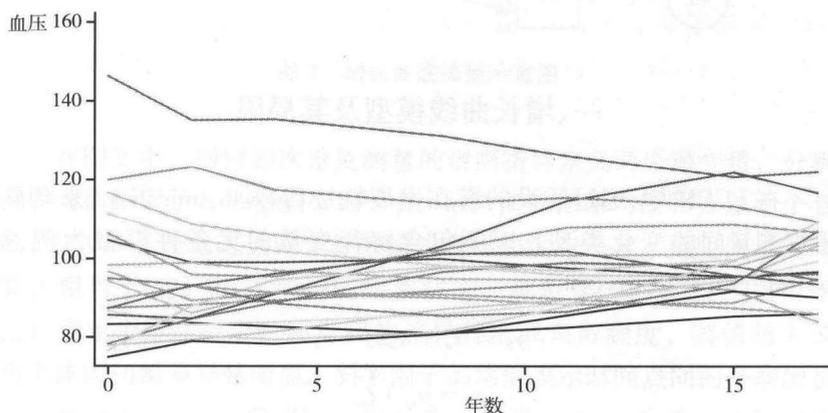
关键词:追踪研究 潜类别增长模型 增长混合模型 Mplus

追踪研究或纵向研究可以揭示个体随时间的变化趋势、个体间差异,且在变量间因果关系推论方面一般要优于截面数据,因此在社会学、心理学和管理学等诸多社会科学领域日益流行(Ferrer & McArdle, 2010; McArdle, 2009; Ployhart & Vandenberg, 2010)。目前,用于处理追踪数据的统计模型有很多,其中社会科学领域使用最多、同时也为广大研究者所熟知的,有潜在增长曲线模型(latent growth curve models, LGCM; Bollen & Curran, 2006)、多水平模型(multilevel modeling, MLM; Bryk & Raudenbush, 1987)和分层线性模型(hierarchical linear modeling, HLM)、随机系数模型(Laird & Ware, 1982)等。^①上述模型尽管名称各异,但统计原理基本相同。比如,在处理追踪数据时,MLM或HLM与LGCM是等价的(Curran, 2003; MacCallum et al., 1997; Raudenbush, 2001)。

* 感谢两位匿名评审专家对本文提出的意见和建议。本研究受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”(项目编号:CBA130124)、广东省高等学校优秀青年教师培养计划(项目编号:Yq2013068)资助。

① 多水平模型框架下纵向研究的介绍可参见宋时歌、陈华珊,2005。

不过,要注意的是,上述模型均假设研究样本存在相同的发展轨迹即内部同质(homogeneity),然而这一假设并非总能满足。如图1所示,尽管大部分个体血压的增长服从“平稳”的发展轨迹,但仍有部分个体呈现“先高后缓慢下降”或“先低后加速增长”的发展轨迹。当研究的样本呈现如图1所示的发展轨迹时,传统增长模型的群体同质性假设受到严重威胁,因为并非全部个体遵从相同的生长趋势。这个时候,传统的生长曲线模型将不再适合,合理的生长模型应该考虑到群体的异质性。



数据来源:中国健康与营养调查(China Health and Nutrition Survey, CHNS)部分样本。

图1 个体血压增长轨迹(非参数平滑化)

潜类别分析(latent class analysis, LCA; Lazarsfeld & Henry, 1968)是基于模型的处理群体异质性的方法。将传统增长模型与潜类别分析相结合的模式,则既可以刻画增长趋势又可以考虑群体异质性。其中,组基增长模型(group-based trajectory modeling; Nagin, 1999, 2005)或潜类增长模型(latent class growth model, LCGM)与增长混合模型(growth mixture modeling, GMM; Muthén & Muthén, 2000; Muthén & Shedden, 1999)是目前两种最常用也是影响最大的处理群体异质增长的模型。二者的区别主要在于类别组内的发展轨迹是否存在方差变异。

GMM作为增长模型的更一般形式逐渐引起国际学术界的注意,应

用研究越来越多,包括心理学(Koss et al., 2013)、教育学(Parrila et al., 2005)、公共卫生(Abroms et al., 2005)、管理学(Qureshi & Fang, 2011)和犯罪学(Morris & Slocum, 2012)等。然而 GMM 在国内尚未见实际使用,希望本文对尚未接触该模型的国内社会科学研究者理解和使用 GMM 有所助益,为纵向数据分析提供一种新的方法选择。本文将从传统增长曲线模型出发,沿着 GMM 的发展路径,从应用性角度依次介绍 LGCM、LCM、LCGM 和 GMM,并通过一个实例简要介绍 3 种增长模型在 Mplus 软件中的实现。最后,结合 GMM 实际使用过程中的常见问题,对最新的进展进行简单评介。

一、增长曲线模型及其局限

在 LGCM 中,通过假设的潜在发展轨迹(growth trajectory)来刻画重复测量间的变化类型。以不包含解释变量的无条件模型为例,^① LGCM 的方程表达式如下:

$$y_{it} = \alpha_{it} + \lambda_{it}\beta_{it} + \varepsilon_{it}$$

$$\alpha_{it} = \mu_{\alpha} + \zeta_{\alpha i}$$

$$\beta_{it} = \mu_{\beta} + \zeta_{\beta i}$$

其中, $i=1, 2, \dots, N$, 表示个体, N 为样本量; $t=1, 2, \dots, T$, 表示测量时间点, T 为测量总次数。 y_{it} 表示个体 i 在时间点 t 时的得分, α_{it} 为个体 i 轨迹的截距即初始水平; β_{it} 为个体 i 成长轨迹的斜率, λ_{it} 表示时间点 t 时的值; ε_{it} 为个体 i 在时间点 t 时的残差项。 μ_{α} 和 μ_{β} 分别表示全部个体截距和斜率的均值即总均值, 由于每个个体具有相同的 μ_{α} 和 μ_{β} , 因此也称固定系数; $\zeta_{\alpha i}$ 和 $\zeta_{\beta i}$ 分别表示个体截距和斜率与对应的总均值间的差异, 每个个体均有一个特定的值, 所以也称作随机系数。

通过路径图可以更好地表达 LGCM 的特点, 包含四次测量的路径图如图 2:

^① 关于 LGCM 更多的内容可以参见 Bollen & Curran, 2006。

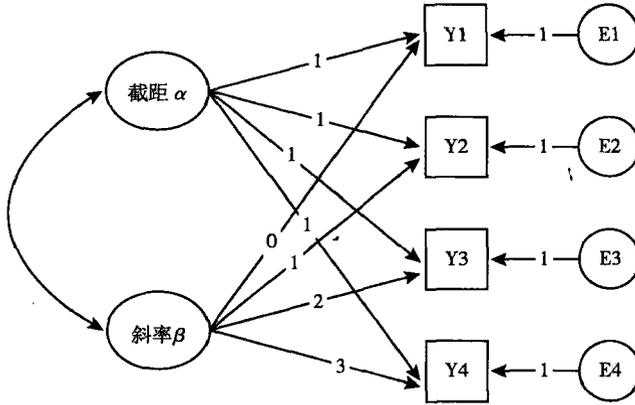


图2 增长曲线模型示意图

在图2中,通过四次重复测量的观测指标定义两个潜变量,分别代表潜截距因子 α 和潜斜率因子 β 。 α 和 β 各存在均值和方差两个参数,LGCM正是通过这两个参数描述组间和组内差异:截距因子的均值(相当于 μ_α)描述平均的初始状态,截距因子的方差(相当于 $\zeta_{\alpha i}$)表示个体在特定时间点间的差异程度或离散程度,其值越大说明个体间初始差异越明显。斜率因子的均值表示时间点间的平均增长率(相当于 μ_β),而斜率因子的方差(相当于 $\zeta_{\beta i}$)反映个体间增长率差异的大小,方差越大表明个体间发展轨迹差异越明显。

重复测量在截距因子上的负荷固定为1,表示每次测量时的截距不变,在斜率因子上的负荷表示时间效应(Meredith & Tisak, 1990),可以设定具体的值,也可以允许自由估计,不同的设定表示不同的时间函数或增长类型。例如,上述四次测量分别为第1次,6个月,12个月和24个月时的测量,可将四次时间设定为等距的间隔 $t_1 = 0$, $t_2 = 6$, $t_3 = 12$ 和 $t_4 = 24$ 或 $t_1 = 0$, $t_2 = 1$, $t_3 = 2$ 和 $t_4 = 4$,表示时间函数为线性即线性发展模型。 α 和 β 之间的双箭头表示两个因子间的相关,用于说明个体变化的截距和变化斜率间的关系。正相关表明较高的初始值对应正向增长斜率;负相关表明,较高的初始值对应负向增长斜率。

在传统的LGCM中,假设群体同质即群体内所有成员享有完全相同或类似的平均增长轨迹(即相同的截距和斜率)。然而在有些情

况下（如图 1 所示），个体间的差异较大，并不一定能满足 LGCM 群体同质的假设前提。此时需要考虑群体可能存在的异质性，即需要引入潜类别变量（categorical latent variable）。

二、引入潜类别概念

潜类别模型是通过间断的潜变量（即潜类别）来解释外显指标间的关联、进而维持其局部独立性的统计模型（Lazarsfeld & Henry, 1968）。其基本假设是，外显变量各种反应的概率分布可以由少数互斥的潜类别变量来解释，每种类别对各外显变量的反应选择都有特定的倾向（邱皓政，2008；Collins & Lanza, 2010）。

LCM 假定任意两个观测指标间的关联可以通过潜变量解释，即达成局部独立性。例如，对于一个潜类别变量 X 和四个观测指标（A、B、C 和 D），LCM 可表达为潜类别概率和条件概率的连乘积：

$$\pi_{ijkl}^{ABCD} = \pi_i^X \pi_{ij}^{A|X} \pi_{jk}^{B|X} \pi_{kl}^{C|X} \pi_l^{D|X}$$

其中， π_i^X 为观测数据归于潜类别变量 X 某一类别的概率， $t =$ 类别个数。 $i、j、k、l$ 分别表示条目 A、B、C 和 D 的选项数目， $\pi_{ij}^{A|X}$ 表示属于某类别 t 的个体在条目 A 的 i 选项上的条件概率，以此类推。

一旦最优的 LCM 拟合成功，就需要将每个个体归入不同的潜类别，亦即确定每个个体的潜类别属性（class membership）。在 LCM 中，采用的分类依据是贝叶斯后验概率（posterior probability），公式如下：

$$1 \quad \pi_{ijkl}^{X|ABCD} = \frac{\pi_{ijkl}^{ABCDX}}{\sum_{t=1}^T \pi_{ijkl}^{ABCDX}}, t = 1, 2, \dots, T$$

后验类别概率是根据个体的作答类型在 LCM 拟合后估计得来，其值表示个体属于某一类别的可能性大小。

目前，LCM 作为处理群体异质性的方法可以和其他模型结合形成更一般的分析模型。例如，与因子模型（factor model）结合形成因子混合模型（factor mixture model, FMM；Lubke & Muthén, 2005；

Muthén, 2008; Muthén & Shedden, 1999); 与增长曲线模型结合即所谓增长混合模型。^①

三、潜类增长模型与增长混合模型

在 LCM 和 LCGM 的基础上, 可将 GMM 和 LCGM 理解为: 先将样本分成若干潜类别组, 然后在每个类别组内建立 LCGM 用于描述类别组内个体随时间变化的情况。不过在 GMM 和 LCGM 中这两步是同时完成的。因此, 在 GMM 和 LCGM 中同时存在两种潜变量: (1) 连续潜变量, 用于描述初始差异和发展趋势的随机截距和随机斜率因子, 同 LCGM; (2) 类别潜变量, 通过将群体分成互斥的潜类别亚组来描述群体的异质性。

GMM 和 LCGM 的方程表达式与 LCGM 类似, 只是将其扩展到包含潜类别变量的形式。

GMM:

$$y_{it} = \sum_{k=1}^k p(c = k) [\alpha_{ik} + \lambda_t \beta_{ik} + \varepsilon_{y_{it}k}]$$

$$\alpha_{ki} = \mu_{\alpha k} + \zeta_{\alpha ik}$$

$$\beta_{ki} = \mu_{\beta k} + \zeta_{\beta ik}$$

LCGM:

$$y_{it} = \sum_{k=1}^k p(c = k) [\alpha_{ik} + \lambda_t \beta_{ik} + \varepsilon_{y_{it}k}]$$

$$\alpha_{ki} = \mu_{\alpha k}$$

$$\beta_{ki} = \mu_{\beta k}$$

其中, C 为类别潜变量, 共有 k 个水平, p 为类别概率, i 为个体, t 为测量时间, μ_{α} 和 μ_{β} 分别表示全部个体截距和斜率的均值即总均值; $\zeta_{\alpha i}$ 和 $\zeta_{\beta i}$ 分别表示个体截距和斜率与对应的总均值间的差异, 每个个体均有一个特定的值。 $\mu_{\alpha k}$, $\mu_{\beta k}$, $\zeta_{\alpha ik}$ 和 $\zeta_{\beta ik}$ 增加了下标 k , 意指

^① 通常将同时包含类别潜变量和连续潜变量的模型称为混合模型 (mixture model)。文献中还存在另外一种混合模型 (mixed model) 或混合效应模型 (mixed effect model), 通常指同时包含固定系数和随机系数的模型 (例如 West et al., 2006)。

类别特定的相应参数。

图3呈现了一例有关酗酒行为研究中的GMM和LCGM的分布形态及增长模式。具体来说，a和b分别表示GMM和LCGM的增长模式图（Muthén, 2008），c和d分别对应某个时点各自分布的截面图（Feldman et al., 2009）。从图中可以非常直观地看出两个模型间的区别。a和b中的粗线代表类别组内群体平均增长曲线，同LCGM。a图中除了粗线外还有代表类别组内个体差异的增长曲线（细线），而b图没有细线表示类别组内不存在个体差异（类别组内方差为0）。

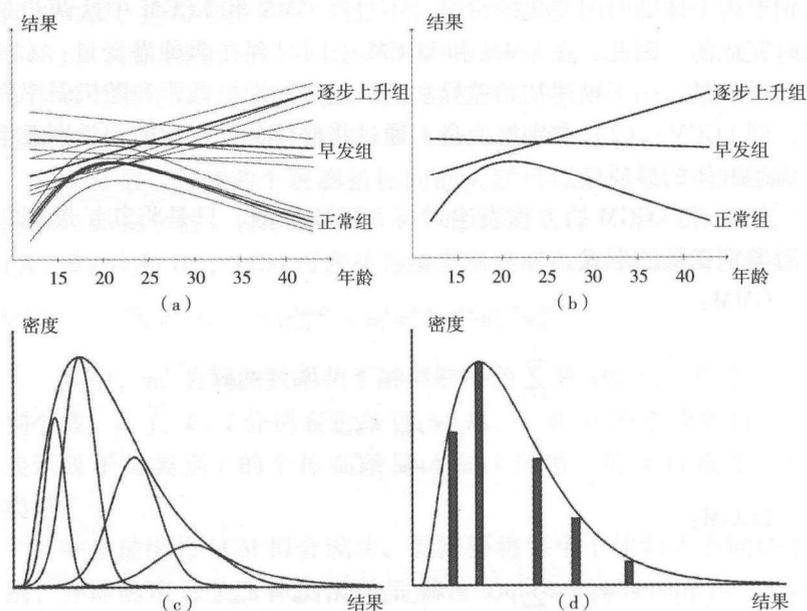


图3 GMM和LCGM示意图

根据上述模型方程表达式及示意图不难发现三者之间的关系：LCGM和LCGM是GMM的特例。LCGM假设所有个体来自同质性群体，所有个体享有相同的平均增长曲线，个体与平均增长曲线的差异（个体间差异）通过增长因子方差来表达。当潜类别变量只有一个水平，即 $k=1$ 时，GMM简化为LCGM。

当类别组内不存在方差变异即将类别组内的增长因子方差固定为0时，单个类别组内所有个体具有完全相同的平均增长曲线，此时称

为 LCGM。GMM 则兼而有之,估计类别组内的平均增长曲线的同时也估计增长因子的方差。

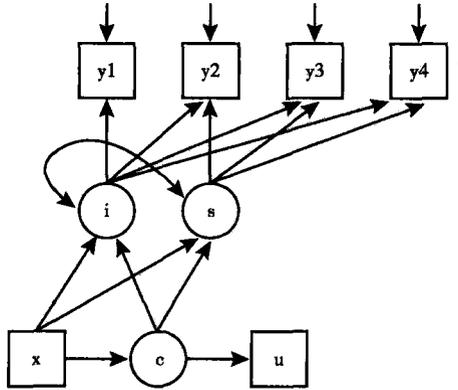


图4 包含协变量和结局变量的 GMM 路径图 (图片来源: Mplus 手册)

研究实践中除了使用增长模型刻画个体发展过程外,还需要考察其他关键变量对此过程产生的影响或受此过程影响的结局变量(outcome variable)。最近的研究发现,包含协变量不仅对探讨变量之间的关系有重要意义,对 GMM 的潜类别个数的确定也同样重要(Li & Hser, 2011; Muthén, 2004; Tofighi & Enders, 2008)。^①图4是一个包含协变量和结局变量的 GMM 示意图。图中 x 为自变量,可以是任何类型(连续、类别或计数型等)的变量,由 x 指向 i 和 s 的路径表示线性回归, x 指向 c 的路径代表多项式回归。潜类别变量 c 指向结局变量 u 的路径依据结局变量的类型可以是相应形式的回归。

四、GMM 建模:一个实例

下面以具体实例,分别演示上述3个模型的分析过程。出于演示的目的,这里主要强调分析过程,而不强调数据结果的理论意义。

^① 不过最近的模拟研究结果并不一致(Li & Hser, 2011; Tofighi & Enders, 2008),所以现在下结论为时尚早。

(一) 数据与建模工具

数据来自美国北卡罗来纳大学和中国疾控中心联合实施的中国健康与营养调查 (CHNS)。该调查从 1989 年开始, 分别于 1991、1993、1997、2000、2004、2006、2009、2011 年进行了追踪调查。^① CHNS 数据样本覆盖大陆 9 个省份, 包含了 4400 户 26000 人。这里选取 1989 年 40 - 50 岁组人群作为分析的对象, 关注的问题是队列 (cohort) 平均血压 $[(\text{收缩压} + 2 \times \text{舒张压}) / 3]$ 随年龄变化的轨迹。CHNS 数据中该队列共有 865 人, 从 1989 - 2006 年共 7 轮数据。^② 7 次测量血压的均值和相关系数矩阵等描述统计量呈现在表 1 中。

表 1 7 次测量血压的描述性统计及相关矩阵

	T89	T91	T93	T97	T00	T04	T06
n	849	761	720	552	571	534	521
均值	88.003	87.837	89.420	92.422	93.988	95.911	96.436
标准差	11.384	12.1786	11.608	13.157	13.133	14.041	14.152
T89	1						
T91	.527	1					
T93	.452	.493	1				
T97	.467	.498	.483	1			
T00	.357	.419	.451	.556	1		
T04	.274	.316	.296	.415	.447	1	
T06	.319	.265	.348	.406	.420	.496	1

建模工具采用的是 Mplus。^③ Mplus 是一款功能强大的潜变量建模软件, 其综合了多个潜变量模型于一个统一的分析框架。Mplus 主要

① 该数据可从项目官网下载 (<http://www.cpc.unc.edu/projects/china>)。

② 影响 GMM 估计准确性的因素有测量次数 (指标个数)、样本量、类别间距 (类别组分布间的距离大小) 以及数据缺失的比例。但目前的研究非常稀少, 一方面是 GMM 刚刚兴起, 另一方面应用学者学习、掌握应用有个时间差。但就目前为数不多的模拟研究来看, 样本量的要求与上述 4 个变量相关。在其他条件不变的情况下, 样本量随测量次数增加而降低, 当到一定程度时影响趋弱。通常一个研究中, 类别间距和数据缺失都很难控制, 样本量和测量次数相对比较容易实现, 所以在样本量不能增加的情况下, 可以通过增加测量次数来实现估计结果的精确性 (Kim, 2012)。

③ 本文示例采用最新的 Mplus7.0 版本。示例所用的带注解的 Mplus 语句及具体设置说明可以向作者索取。

采用语句命令,但语法非常简洁,特别是在处理复杂模型时更是非常方便和简洁(关于此软件更进一步的介绍,参见 Muthén & Muthén, 2012; 王孟成, 2014)。

(二) 建模过程及一般原则

总的来说, GMM 作为潜变量模型中的一员其建模过程遵循一般潜变量建模的通用流程(王孟成, 2014),但在一些细节处存在不同。

1. 研究假设

首先,要明确选择 GMM 作为分析方法的依据。例如,是 LGCM 模型拟合不理想,还是从研究目的和理论预期考虑。

同 LCM 一样(邱皓政, 2008; Collins & Lanza, 2010), GMM 模型的设定从基线模型开始。先设定单类别增长模型,然后逐渐增加类别个数。理想的情况是,研究根据理论预测设定潜类别数目,但实际应用上多采用探索的思路,根据数据拟合结果选择模型。

2. 模型估计

参数估计具体采用 EM 算法。一般来说, LCM 的参数估计有两种基于极大似然估计的迭代算法:期望最大(expectation-maximization, EM)和牛顿-拉夫森算法(Newton-Raphson)。迭代算法的一般过程分成两个阶段:第一阶段,以一个(组)开始值(starting values)为初始估计值进行估计以获得最大值;第二阶段,用第一阶段的估计最大值再进行估计,直到达到设定的聚合标准。^①

GMM 的估计方法依据观测指标的类型而定。传统上, GMM 的指标为连续变量,但在更宽泛的意义上来说(Mplus 下的一般潜变量模型框架), GMM 可以构建任何尺度变量的模型。^②就 Mplus 软件来说,当观测变量为连续变量时,默认的参数估计方法为稳健较大似然

① 两种方法各有优缺点,但两种算法均易产生局部最大化解(local maxima)而非总体最大化解(global maxima)。解决的途径是,估计同一模型时设置不同的开始值,如果结果差异较大说明所获得的解很可能是局部解。在某些统计软件中,如 Mplus,程序默认第一阶段从 10 个随机初始值开始估计,第二阶段,使用第一阶段获得的 2 个最大值进行估计。研究者根据研究需要修改默认设置,通常的设置第一阶段设置 200-500 个初始值,第二阶段设置 20-50 个初始值。

② 通常也称作一般增长混合模型(general growth mixture modeling, GGMM),本文主要涉及连续结局变量。

估计 (MLR)。除此之外, Mplus 还支持贝叶斯估计, 该方法在最近的模拟研究中表现不俗 (Sarah, 2013)。

3. 模型解释与评价

在确定潜类别个数后, 需要对模型的理论意义做出解释。或者引入协变量对不同类别增长趋势进行预测。

一般来讲, 模型评价遵循两条原则: 实际意义和统计指标。实际意义的评价应结合理论、前人的研究结果和结果的可解释性。

混合模型拟合评价的指标主要有 Pearson 卡方检验和似然比卡方 G^2 (LL) 检验, 以及信息评价指标 AIC、BIC 和样本校正的 BIC (sample size-adjusted BIC, aBIC), 这几种统计量都是通过比较期望值与实际值差异来判断拟合优劣, 统计数值越小表示拟合得越好。然而, 当样本量很大时, 卡方统计检验变得十分保守, 即使期望值与观察值相差不大, 也很容易判断为差异显著。所以大部分的实证研究都使用 BIC 指标作为模型适配度比较的指标, 一般是选择 BIC 最小的模型作为最佳模型。

混合模型的一个主要特点是将异质性群体分类, 因此在评价混合模型时, 常常使用 Entropy 去评价分类精确性, 公式如下:

$$E_k = 1 - \frac{\sum_i \sum_k (-p_{ik} \ln p_{ik})}{n \ln K}$$

P_{ik} 为个体 i 属于类别 k 的后验概率。其取值范围在 0 - 1 之间, 越接近 1 表明分类越精确。有学者 (Lubke & Muthén, 2007) 指出, Entropy < 0.60 相当于超过 20% 的个体存在分类错误; Entropy \geq 0.80 表明分类准确率超过 90%。由于模拟条件的限制, 这一标准只能作为参考。

4. 模型比较

嵌套模型可以采用似然比检验 (likelihood ratio test, LRT), 然而不同潜类别个数的混合模型并非彼此嵌套的关系 (Nylund et al., 2007), 主要由于 $k-1$ 个类别是 k 个类别模型的特例, 在 k 个类别模型中, 第一个类别的概率被设定为 0, 导致差异间不再呈卡方分布。替代的检验统计量主要有基于 Bootstrap 的似然比检验 (BLRT; McLachlan & Peel, 2000) 和 LMR (Lo-Mendell-Rubin, LMR; Lo et al., 2001) 似然比检验。

BLRT 使用 Bootstrap 抽样估计两个嵌套模型间的对数似然比差异分布。BLRT 主要比较 $k-1$ 个和 k 个类别模型间的拟合差异。例如, 对于一个有 4 个类别的 LCA 模型, BLRT 的 p 值比较 3 个类别和 4 个类别模型间拟合的差异。显著的 BLRT 的 p 值表示 4 个类别的模型比 3 个类别的模型拟合显著改善。不显著的 BLRT 的 p 值则表明 4 个类别的模型并未比三个类别模型显著改善拟合。LMR^① 与 BLRT 类似, 用于比较 $k-1$ 个和 k 个类别模型间的拟合差异。显著的 LMR 的 p 值表明, k 个类别模型优于 $k-1$ 个类别模型。

(三) 模型分析示例

1. LGCM

这里估计的 LGCM 模型为自由估计时间参数模型, 即设定前三次的时间分, 其他时间点的负荷由数据估计。下述 GMM 和 LCGM 的增长均采用此设置。该模型的拟合指数结果如下: 卡方值 = 49.728, $df=19$, $p < 0.01$, TLI = 0.973, CFI = 0.976, AIC = 34406.242, BIC = 34482.446, aBIC = 34431.634, RMSEA = 0.043。根据通常的拟合指数评价标准判断, 模型拟合非常好。

截距因子和斜率因子的方差估计分别为 75.805 和 1.634, 均在 0.01 水平上显著, 说明个体在初始水平及增长速度上存在显著个体间差异。截距增长因子和斜率增长因子的相关系数为 -0.348 ($p < 0.01$), 提示个体增长变化率与初始状态存在显著的反向关系, 即初始状态得分高的个体下降速度较慢。

2. LCGM

依据混合模型建模的一般步骤, 在 LGCM 模型的基础上依次提取了 1-6 个类别的 LCGM 模型, 拟合结果呈现在表 2 中。

表 2 的拟合指数提示, 信息指数随类别数量的增加而逐渐减少, 所以很难依据信息指数来选择最佳模型。在此情况下可以采用碎石图来选择 (Petras & Masyn, 2010), BIC 的碎石图结果呈现在图 5 上。仅从图上来看, 在 3 处存在明显的拐点, 因此选取 3 个类别较为合适。

LMR 和 BLRT 的结果并不一致。在所有模型中, BLRT 均提示显

① Mplus 除了提供 LMR 还提供 LMR 的校正值, 只是比较 $k-1$ 的模型为删除第一个类别的模型 (Muthén & Muthén, 2012)。

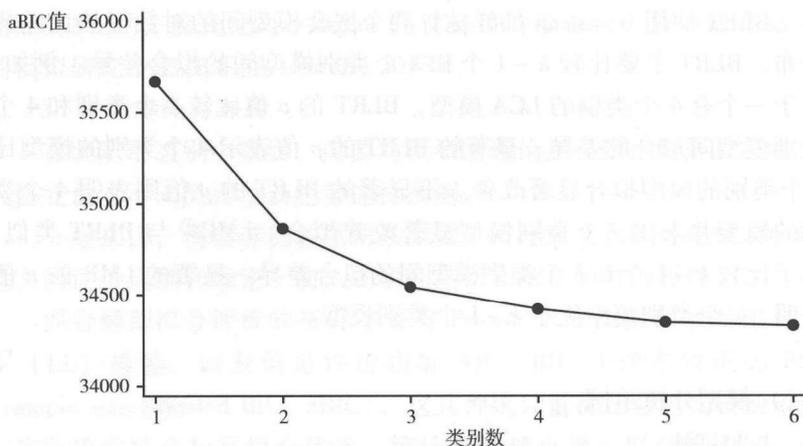


图5 aBIC值的陡坡图

著，而 LMR 在取 4 个类别时 $p = 0.492$ ，提示保留 3 个类别较优。但随着类别数目增加到 5 个时，其结果出现了边缘显著 $p = 0.030$ ，而此时的类别概率差异较大，其中最小的为 0.025，对应的 $N = 20$ ，综合多方信息最后选择保留 3 个类别的模型。三个类别 LCGM 对应的增长因子估计均值见图 6。

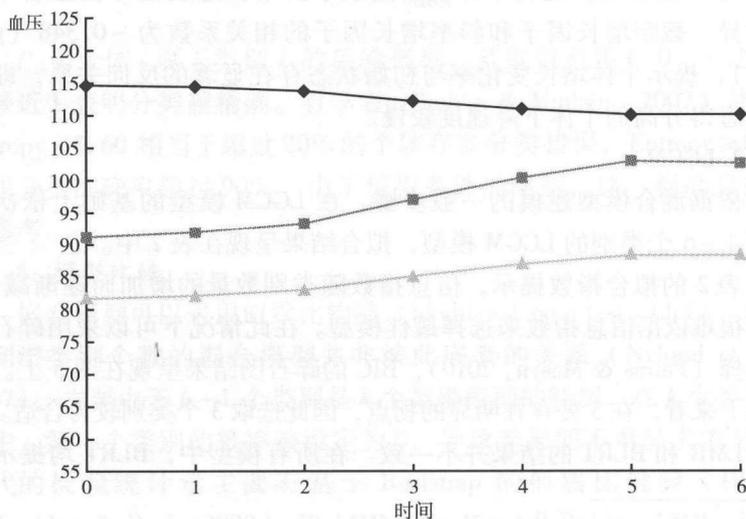


图6 LCGM (k=3) 类别增长轨迹图

如图 6 所示,三个类别组内的个体享有完全相同增长轨迹,并呈现出不同的增长模式。类别组一(图中最上方一组)47人(占全体的 5.4%),截距和斜率的均值分别为 114.575 ($p < 0.01$)和 -0.308 ($p = 0.268$),该组的血压起始值最高且呈逐渐下降趋势,将其命名为“高压缓慢下降组”。类别组二(图中中间的一组)419人,截距和斜率的均值分别为 91.065 ($p < 0.01$)和 0.786 ($p < 0.01$),占 48.4%,该群体初始血压正常随年龄增长而增加,命名为“缓慢增长组”;类别组三(图中最下方的一组)399人,占 46.1%,截距和斜率的均值分别为 81.655 ($p < 0.01$)和 0.452 ($p < 0.01$),命名为“正常组”。

3. GMM

LCCM 作为 GMM 的特例,限定增长因子方差和协方差为 0,即假设类别内所有个体具有完全相同的生长轨迹。这一假设比较苛刻,在本例中即使 LCCM 考虑的群体异质性,增长模型的拟合并没有优化 ($\Delta\text{BIC} = \text{aBIC}_{\text{LCCM}} - \text{aBIC}_{\text{LCCM}} = -111$),下面拟合释放了这一限定的 GMM。由于存在多个类别组,各类别组内的方差和协方差可以设定相等即跨类别组等同(类似测量等值),也可以设定自由估计,限于篇幅下面只报告设定跨组等同的 GMM 结果。

依次抽取 4 个类别 GMM 的拟合结果整理在表 2 中。LMR 和 BLRT 的结论与信息指数的结论并不一致。LMR 和 BLRT 的结果一致提示,选取 3 个类别的结果优于 4 个类别,但信息指数则随类别数量的增加而减少。取三个类别时的分类精确性指标 $\text{Entropy} = 0.871$ 优于 4 个类别的结果 0.849。同时,类别概率也较合理,所以综合来看选取 3 类别模型。

与 LCCM 相比,GMM 各项拟合结果均有显著优化,其中 $\Delta\text{AIC} = 275.348$, $\Delta\text{aBIC} = 200.721$, Entropy 值也从 0.738 增加至 0.871,分类概率也发生变化。

三个类别 GMM 增长趋势见图 7,对应的截距和斜率因子估计值见表 3。粗线表示模型估计均值,细线表示模型估计的个体值。对比图 6 和图 7 不难发现,两个模型估计的增长模式存在相似之处,但差异更为明显。具体来说,类别组 1(图中左下至右上方向一组)起始水平较低,逐步升高,命名为“患病组”,共 34 人,占全体的 3.93%,截距和斜率因子的均值分别为 84.857 ($p < 0.01$)和 2.385 ($p < 0.01$)。

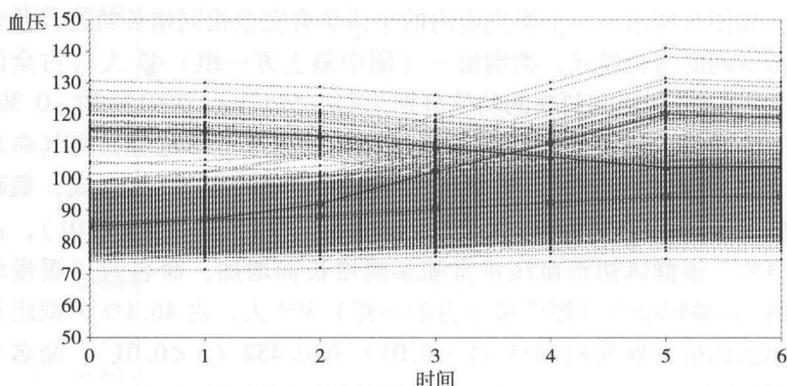


图7 三个类别 GMM 增长趋势图 (估计均值和估计个体值)

类别组 2 (图中左上至右下方向一组), 开始水平较高, 随时间变化逐渐下降, 命名为“康复组”, 共 40 人, 占群体的 4.62%。该组截距和斜率因子的均值分别为 115.844 ($p < 0.01$) 和 -0.831 ($p = 0.021$), 说明该组个体血压初始值存在差异。

类别 3 (图中最下方一组) 始终保持平稳, 命名为“正常组”, 占全体人数的绝大部分 (91.45%), 截距和斜率因子的均值分别为 86.461 ($p < 0.01$) 和 0.535 ($p < 0.01$)。

截距增长因子和斜率增长因子间的协方差 (被限定跨类别等值) 为 0.461 ($p = 0.213$), 说明初始水平与增长率之间关系不大。截距因子的方差为 36.011 ($p < 0.01$), 说明各类别组内的个体间血压初始值存在差异。而斜率因子的方差为 0.051 ($p = 0.293$), 说明各类别组内个体间血压增长率差异不显著。

4. 带有协变量的 GMM

在实践中可以纳入协变量 (如人口学变量) 进行混合模型分析, 以提高分类精确性 (Petras & Masyn, 2010; Nylund et al., 2007)。下面以性别 (其中 0 代表女性, 1 代表男性) 作为预测协变量演示带有协变量的 GMM。

从表 2 的拟合结果来看, 信息指数支持保留 3 个类别, 而 LMR 支持选择 2 个类别, 此时的分类精确性也较高。出于与无条件模型比较的目的, 这里选择报告 3 个类别的结果, 模型估计的参数见表 3。

表 2 LCGM 和 GMM 模型拟合信息汇总

模型	k	C2/LL	AIC	BIC	aBIC	Entropy	LMR	BLRT	类别概率
LCGM									
1C	13	-17810.415	35646.829	35708.745	35667.460	—	—	—	1
2C	16	-17403.802	34839.605	34915.808	34864.996	.753	<.001	<.001	.244/.756
3C	19	-17237.242	34512.484	34602.976	34542.637	.738	.004	<.0001	.054/.484/.461
4C	22	-17171.30	34386.604	34491.384	34421.517	.704	.492	<.0001	.571/.260/.045/.124
5C	25	-17129.434	34308.869	34427.937	34348.544	.729	.030	<.0001	.025/.259/.060/.128/.527
6C	28	-17114.184	34284.368	34417.724	34328.803	.698	.624	<.0001	.476/.047/.204/.212/.015/.046
GMM									
1C	16	-17187.121	34406.242	34482.446	34431.634	—	—	—	1
2C	19	-17118.895	34275.790	34366.282	34305.943	.947	<.001	<.001	.954/.046
3C	22	-17096.568	34237.136	34341.916	34341.916	.871	.021	<.001	.039/.047/.914
4C	25	-17092.110	34234.220	34353.288	34273.895	.849	.506	.666	.047/.039/.898/.015
GMM - 协变量									
2C	20	-17117.755	34275.510	34370.765	34307.250	.948	<.001	<.001	.9534/.046
3C	26	-17082.996	34217.993	34341.824	34259.255	.873	.068	<.000	.914/.050/.036

注:拟合最优的模型被加粗。

从模型分类概率来看,加入协变量后,3个类别的比例有所变化,说明协变量的加入确实会影响模型的结果,而且对类别数目的保留也有所影响。具体到本例来说,加入协变量后,取三个类别模型的LMR的值由0.021增加到0.068,使得LMR更倾向于选择2个类别模型。对各类别增长因子的均值和方差也有一定程度的影响(见表3)。

表3 GMM 各类别截距和斜率因子估计均值

	n	截距	斜率
无条件 GMM			
类别 1	34	84.857 (2.139)	2.385 (.617)
类别 2	791	86.461 (.388)	.535 (.133)
类别 3	40	115.844 (2.878)	-.831 (.361)
有条件 GMM			
类别 1	791	84.856 (.504)	.558 (.147)
类别 2	31	83.592 (2.027)	2.284 (.621)
类别 3	43	113.944 (2.641)	-.551 (.281)

注:括号内为标准误。

五、结论与讨论

追踪研究或纵向研究在社会学、心理学和管理学等诸多社会科学领域日益流行(Ferrer & McArdle, 2010; McArdle, 2009; Ployhart & Vandenberg, 2010),同时用于处理追踪数据的统计模型也日新月异。传统的增长曲线模型假设所有的个体遵循相同的发展轨迹,当存在不同发展轨迹时,这一假定不切实际。更合理的模型应该考虑增长的异质性。

本文在传统LGCM的基础上,介绍了将LGCM与LCM相结合的更为一般的GMM,以及几种常见的变式。GMM作为新近发展的潜类别增长模型逐渐引起国际学术界的注意,应用研究和方法学研究越来越多(Hix-Small et al., 2004; Kim, 2012; Kim & Kim, 2012; Li & Hser, 2011; Muthén & Muthén, 2000),但在国内尚未见实际使用,希望本文的介绍能为国内社会科学研究者提供一种新的方法选择。

作为一个不断发展的新方法,人们对 GMM 的认识也在不断加深。总的来说,当前混合模型(包括 LCA、GMM)使用过程中最常见也是最容易让应用研究者困惑的问题主要存在如下几个:

首先,当采用混合模型拟合数据时发现群体异质性即提取二个以上潜类别组时,结果如何解释。混合模型(mixture model)的分析基础是有限混合模型(finite mixture model)。有限混合模型假设任意总体分布都可以由多个正态亚分布混合而成,或者说,任意分布的总体都可以分解成多个正态分布的分样本。通常,非正态的聚合分布(aggregate distribution)是混合模型能抽取多个潜类别组的关键(Bauer & Curran, 2003),然而对非正态组合分布存在两种不同的理论解释。第一,样本非正态分布是由于不同质的个案混合而成的,即存在潜类别组或群体异质性。从研究样本中抽取的多个潜类别组正好说明样本中确实存在群体异质性。第二,由于抽样误差的影响,即使从正态分布的总体中取样,也可能得到非正态分布的样本。此时的潜类别组则是由于抽样误差造成的,而不是群体异质性的体现。

鲍尔和科伦(Bauer & Curran, 2003)通过模拟证实了上述两种情况在实际应用中给研究者产生的困扰。他们模拟实际只存在单个潜类别的数据,但存在不同程度的非正态性。结果发现,如果非正态存在,即使本身只存在单个潜类别组,实际分析过程中也可以提取出2个(或多个)潜类别组。这一研究结果说明,数据非正态会导致高估实际存在潜类别数目。

在实际应用中,研究者如何区分所得结果是真实的群体异质性还是抽样误差造成的虚假异质性?该问题涉及如何正确地探测真实群体异质性和排除虚假的异质性。正确探测群体异质性或潜类别个数保留可以从统计指标和实际的理论预测两个方面入手。保留潜类别个数的统计指标将在下面讨论,理论预测将是解释群体异质性结果最重要的途径(Bauer & Curran, 2003; Muthén, 2003)。由于小样本容量更容易扭曲总体分布形态,所以扩大样本量是避免虚假异质性的有效手段之一。

其次,保留正确的潜类别个数是探测真实群体异质性的关键,也是目前混合模型领域的热点和难点问题。高估或低估潜类别个数都将影响结果推论的准确性,不少模拟研究对这一问题进行了探索(Nylund et al., 2007; Yang, 2006)。

杨 (Yang, 2006) 的模拟研究发现, aBIC 是分类准确度最高的信息指数, 其前提是每个类别至少要有 50 个被试。该研究还发现, AIC 在确定类别个数时表现欠佳 (也见 Nylund et al., 2007)。尼隆德等 (Nylund et al., 2007) 的研究在多种混合模型 (LCA, 因子混合模型和 GMM) 中比较了 BLRT、LMR 和信息指数, 结果发现 BLRT 和 BIC 分别为基于似然比和信息指数中表现最好的指标, 而 BLRT 又优于 BIC, 且在多种情况下 (样本量、模型类型、指标个数及潜类别个数) 具有一致的表现。

在实际应用中, 各评价指标之间往往存在不一致。例如, BLRT 的 p 值显著, 而 LMR 的 p 值则远大于 0。如果遇到此种情况, 应结合分类的实际意义和类别包含样本数来确定最终的类别数目。具体来说, 即使各项指标提示保留 m 个类别, 而其中的一个类别个体数目有限, 或者不易解释时, 应该考虑 $m-1$ 模型。

通常似然值或信息指数会随着类别数目的增加而减少, 但有时类别数目增加了很多也未获得最佳拟合模型。例如, BIC 值随类别数单调递减, 始终未见最低值。针对这种情况, 可以采用类似探索性因子分析确定因子个数的陡坡图检验 (Petras & Masyn, 2010)。例如, 图 5 就是根据表 2 中的 aBIC 值从高到低依次排列的陡坡图。仅从图上来看, 在 3 处存在明显的拐点, 因此选取 3 个类别是合适的。该方法存在的最大不足是若陡坡图未出现明显的拐点, 则不宜使用此法。

最后, 需要指出的是, 尽管 GMM 近年来引起多方关注, 但对 GMM 的认识并没有像其他潜变量建模方法那样深入全面, 在将来的研究中, 方法学者应该着重关注如下方面的问题: 第一, 全面探讨缺失率和缺失类型对 GMM 结果 (如, 潜类别个数保留、参数估计) 可靠性的影响。尽管有个别方法学研究涉及了缺失值问题 (Kim, 2012), 然而尚未有研究全面系统探讨不同缺失率及缺失类型对 GMM 收敛、潜类别数量保留等的影响。第二, 各种潜类别个数保留指数在不同建模条件下的表现。与结构方程模型的这种拟合指数表现的研究相比, 当前对各种 GMM 类别保留指数表现的研究相当稀少, 而且结果也不一致 (Peugh & Fan, 2012; Nylund et al., 2007; Tofighi & Enders, 2008; Yang, 2006)。第三, 对 GMM 结果精确性影响的其他方面。例如样本量 (Kim, 2012) 与是否包含协变量 (Li & Hser, 2011)。

总之, GMM 作为潜类别模型和增长曲线模型更一般的数据建模方法在分析追踪数据发展趋势的同时兼顾群体异质性, 在短短的十几年里得到研究者的广泛关注。随着统计分析软件的普及, GMM 将会为更多的研究者应用。

参考文献:

- 邱皓政, 2008, 《潜在类别模型的原理与技术》, 北京: 教育科学出版社。
- 宋时歌、陈华珊, 2005, 《纵贯性数据与生长模型在社会科学实证研究中的应用》, 《社会学研究》第 5 期。
- 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。
- Abrams, L., B. Simons-Morton, D. L. Haynie, & R. Chen, 2005, "Psychosocial Predictors of Smoking Trajectories During Middle and High School." *Addiction* 100.
- Bauer, D. J. & P. J. Curran 2003, "Distributional Assumptions of Growth Mixture Models: Implications for Overextraction of Latent Trajectory Classes." *Psychological Methods* 8.
- 2004, "The Integration of Continuous and Discrete Latent Variable Models: Potential Problems and Promising Opportunities." *Psychological Methods* 9.
- Bollen, K. A. & P. J. Curran 2006, *Latent Curve Models: A Structural Equation Perspective*. Hoboken, NJ: Wiley.
- Bryk, A. S. & S. W. Raudenbush 1987, "Application of Hierarchical Linear Models to Assessing Change." *Psychological Bulletin* 101.
- Collins, L. M. & S. T. Lanza 2010, *Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences*. New York: Wiley.
- Curran, P. J. 2003, "Have Multilevel Models been Structural Equation Models all Along?" *Multivariate Behavioral Research* 38.
- Feldman, B. J., K. E. Masyn & R. D. Conger 2009, "New Approaches to Studying Problem Behaviors: A Comparison of Methods for Modeling Longitudinal, Categorical Adolescent Drinking Data." *Developmental Psychology* 45.
- Ferrer, E. & J. J. McArdle 2010, "Longitudinal Modeling of Developmental Changes in Psychological Research." *Current Directions in Psychological Science* 19.
- Hipp, J. & D. J. Bauer 2006, "Local Solutions in the Estimation of Growth Mixture Models." *Psychological Methods* 11.
- Hix-Small, H., T. E. Duncan, S. C. Duncan & H. Okut 2004, "A Multivariate Associative Finite Growth Mixture Modeling Approach Examining Adolescent Alcohol and Marijuana Use." *Journal of Psychopathology and Behavioral Assessment* 26.
- Kim, S-Y. 2012, "Sample Size Requirements in Single-and Multiphase Growth Mixture Models: A Monte Carlo Simulation Study." *Structural Equation Modeling* 19.
- Kim, S-Y. & J.-S. Kim 2012, "Investigating Stage-sequential Growth Mixture Models with Multiphase Longitudinal Data." *Structural Equation Modeling* 19.

- Koss, K. J., M. R. W. George, P. T. Davies, D. Cicchetti, E. M. Cummings & M. L. Sturge-Apple 2013, "Patterns of Children's Adrenocortical Reactivity to Interparental Conflict and Associations with Child Adjustment: A Growth Mixture Modeling Approach." *Developmental Psychology* 49.
- Laird, N. M. & J. H. Ware 1982, "Random Effects Models for Longitudinal Data." *Biometrics* 38.
- Lazarsfeld, P. F. & N. W. Henry, 1968, *Latent Structure Analysis*. Boston, MA: Houghton Mifflin.
- Li, L. & Yih-Ing Hser 2011, "On Inclusion of Covariates for Class Enumeration of Growth Mixture Models." *Multivariate Behavioral Research* 46.
- Lo, Y., N. Mendell & D. B. Rubin, 2001, "Testing the Number of Components in A Normal Mixture." *Biometrika* 88.
- Lubke, G. H. & B. Muthén, 2005, "Investigating Population Heterogeneity with Factor Mixture Models." *Psychological Methods* 10.
- 2007, "Performance of Factor Mixture Models as a Function of Covariate Effects, Model Size, and Class-specific Parameters." *Structural Equation Modeling* 14.
- MacCallum, R. C., C. Kim, W. B. Malarkey & J. K. Kiecolt-Glaser 1997, "Studying Multivariate Change Using Multilevel Models and Latent Curve Models." *Multivariate Behavioral Research* 32.
- McArdle, J. J. 2009, "Latent Variable Modeling of Differences and Changes with Longitudinal Data." *Annual Review of Psychology* 60.
- McLachlan, G. J. & D. Peel 2000, *Finite Mixture Models*. New York: Wiley.
- Meredith, W. & J. Tisak 1990, "Latent Curve Analysis." *Psychometrika* 55.
- Morris, N. A. & L. A. Slocum 2012, "Estimating Country-Level Terrorism Trends Using Group-Based Trajectory Analyses: Latent Class Growth Analysis and General Mixture Modeling." *Journal of Quantitative Criminology* 28.
- Muthén, B. O. 2003, "Statistical and Substantive Checking in Growth Mixture Modeling: Comment on Bauer and Curran (2003)." *Psychological Methods* 8.
- 2004, "Latent Variable Analysis: Growth Mixture Modeling and Related Techniques for Longitudinal Data." In D. Kaplan (ed.), *Handbook of Quantitative Methodology for the Social Sciences*. Newbury Park, CA: Sage Publications.
- 2008, "Latent Variable Hybrids." In G. R. Hancock & K. M. Samuelson (eds.), *Advances in Latent Variable Mixture Models*. Charlotte, NC: Information Age Publishing, Inc.
- Muthén, B. O. & L. K. Muthén 2000, "Integrating Person-centered and Variable-centered Analyses: Growth Mixture Modeling with Latent Trajectory Classes." *Alcoholism: Clinical and Experimental Research* 24.
- Muthén, L. K. & B. O. Muthén 2012, *Mplus User's Guide (7th ed.)*. Los Angeles: Muthén & Muthén.
- Muthén, B. & K. Shedden, 1999, "Finite Mixture Modeling with Mixture Outcomes Using the EM

- Algorithm." *Biometrics* 55.
- Nagin, D. S. 1999, "Analyzing Developmental Trajectories: A Semiparametric, Group-based Approach." *Psychological Methods* 4.
- 2005, *Group-Based Modeling of Development Over the Life Course*. Cambridge, MA: Harvard University Press.
- Nylund, K. L., T. Asparouhov & B. O. Muthén 2007, "Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study." *Structural Equation Modeling* 14.
- Parila, R, K. Aunola, E. Leskinen, J. Nurmi & J. R. Kirby 2005, "Development of Individual Differences in Reading: Results from Longitudinal Studies in English and Finnish." *Journal of Educational Psychology* 97.
- Petras, H. & K. Masyn 2010, "General Growth Mixture Analysis with Antecedents and Consequences of Change." In R. Piquero & D. Weisburd (eds.), *Handbook of Quantitative Criminology*. New York, NY: Springer.
- Peugh, J. & X. Fan 2012, "How Well Does Growth Mixture Modeling Identify Heterogeneous Growth Trajectories? A Simulation Study Examining GMM's Performance Characteristics." *Structural Equation Modeling* 19.
- Ployhart, R. E. & R. J. Vandenberg 2010, "Longitudinal Research: The Theory, Design, and Analysis of Change." *Journal of Management* 36.
- Qureshi, I. & Y. Fang 2011, "Socialization in Open Source Software Projects: A Growth Mixture Modeling Approach." *Organizational Research Methods* 14.
- Raudenbush, S. W. 2001, "Comparing Personal Trajectories and Drawing Causal Inferences from Longitudinal Data." *Annual Review of Psychology* 50.
- Sarah, D. 2013, "Mixture Class Recovery in GMM under Varying Degrees of Class Separation: Frequentist versus Bayesian Estimation." *Psychological Methods* 18.
- Tofghi, D. & C. Enders 2008, "Identifying the Correct Number of Classes in Growth Mixture Models." In G. R. Hancock & K. M. Samuelson (eds.), *Advances in Latent Variable Mixture Models*. Charlotte, NC: Information Age.
- West, B., K. B. Welch & A. T. Galecki 2006, "Linear Mixed Models: A Practical Guide Using Statistical Software." London, UK: Chapman and Hall/CRC Press.
- Yang, C. 2006, "Evaluating Latent Class Analyses in Qualitative Phenotype Identification." *Computational Statistics and Data Analysis* 50.

作者单位: 广州大学心理学系 (王孟成)
 中国政法大学社会学院 (毕向阳)
 广州大学心理与脑科学研究中心 (叶浩生)
 责任编辑: 闻翔

Exploration of Chinese Voluntary Service Based on “the Emic Appeals” : A case study of migrant children

..... *Fu Xiaoxing, Liu Shang & Chen Yupei* 196

Abstract: Different from the voluntary services in western countries, whose practices are institutionalized and are based on mature civil societies, the burgeoning voluntary services in China is in the process of development and exploration. In response to current problems such as lack of corresponding institution and poor management, anthropology emphasizes the respect towards service subjects, and suggests volunteers to conduct theoretical construction and application with them. Taking migrant children as an example, this paper has developed an original model of voluntary service based on a three-year field work. The pivotal concept of this model is “the emic appeals”, on the basis of which the authors have created “progressive” strategies in response to appeals of migrant children. Furthermore, the authors have adopted “triangulation method” to ensure the credibility and sustainable development of “the emic appeals”. Through this practical anthropology, the authors hope to provide a valuable methodological inspiration for other disciplines to participate in voluntary service in China.

REVIEW

Growth Mixture Modeling: A method for describing specific class growth trajectory *Wang Mengcheng, Bi Xiangyang & Ye Haosheng* 220

Abstract: Longitudinal research has attracted more and more attention among social science researchers in recent years, because it could describe the within-person change and between-person differences in change simultaneously, along with special advantages in explaining causality. For conventional latent growth curve modeling (LGCM), it is not to be hold to assume all individuals are drawn from a homogeneous population in many applied research situations. When the population is heterogeneous, the growth mixture modeling (GMM) and latent class growth modeling (LCGM; or group-based growth modeling) are the best choices. In fact, LGCM and LCGM are the special types of GMM as the universal model. The purpose of this paper is to provide a non-technical overview of these growth modeling (LGCM, LCGM and GMM). The modeling procedures are illustrated with a practical example. Finally, current debates and issues in the modeling process, as well as the new directions for future research, are briefly discussed.