· 计算机应用 ·

基于分组的轨迹模型及其 SAS 实现*

天津医科大学公共卫生学院流行病与卫生统计学教研室　天津市环境营养与人群健康重点实验室(300070)
宋福曼　王思婷　乔亲群　刘媛媛　陈佳庚　马　骏△

【提　要】 目的　对基于分组的轨迹模型(group-based trajectory modeling, GBTM)的基本原理、适用性及 SAS 软件程序进行介绍。**方法**　使用 SAS 软件中的 PROC TRAJ 过程步,构建不同数据类型的 GBTM 模型。**结果**　选取多项前瞻性数据,根据数据类型构建了 CNORM、ZIP 等单轨迹及双轨迹模型,在此基础上将协变量纳入模型,观察影响发展轨迹的因素。**结论**　GBTM 为解决纵向数据复杂性和结果可解释性之间的平衡问题提供了有价值的工具,同时 SAS 软件可以很好地实现轨迹模型的构建,并为模型的不断拓展提供了较为方便的实现形式。

【关键词】 基于分组的轨迹模型　轨迹分析　纵向数据
【中图分类号】 R195.1　　**【文献标识码】** A　　**DOI** 10.11783/j.issn.1002-3674.2024.03.029

基于分组的轨迹模型(group-based trajectory modeling, GBTM),是一种用于识别随着年龄或时间的推移,某些行为具有相似发展轨迹的个体集群的统计分析方法[1]。经典的发展轨迹分析方法,如潜在曲线分析,大多旨在解释总体平均轨迹中的个体差异,往往忽略了群体中可能存在遵循独特发展轨迹的亚组。而 GBTM 假定总体存在异质性,进而探求总体中各亚组的发展轨迹,研究发展轨迹与影响因素或结果之间的联系。一方面可以充分利用纵向数据;另一方面可以将结果以易于解释的图形或表格等形式呈现。近年来,GBTM 被广泛应用于医学研究,如了解疾病的病因和发展过程[2-4]、识别影响轨迹的因素或进行预测分析[5]、以及观察性研究中的因果推断等[6]。

实际应用中,GBTM 的拟合过程可以通过 SAS、Stata、R 等软件实现。其中 SAS 软件中的 PROC TRAJ 过程步是一种对用户友好的有限混合模型程序。相较其他软件,它的输出结果更丰富,不仅可以直接输出各参数检验结果以及 BIC 值,而且可以在输出数据集中获得后验概率,进而求出平均后验概率(average posterior probability, AvePP)等模型拟合效果指标[7],因此应用较为广泛。

本研究通过构建不同数据类型的轨迹模型,对 GBTM 的基本原理、适用性及相应的 SAS 软件程序进行介绍。

原理简介

1. 模型构建

GBTM 是有限混合模型的专门应用,模型的参数估计值是极大似然估计法(maximum likelihood estimate, MLE)的产物。似然函数的具体形式取决于所分析的数据类型。设 $Y_i=\{y_{i1},y_{i2},y_{i3},\cdots,y_{iT}\}$,用来表示在 T 周期内个体 i 的纵向测量序列。令 $P(Y_i)$ 表示 Y_i 的概率,假设总体中包含 j 个潜在轨迹组,则:

$$P(Y_i)=\sum_{j}^{J}\pi_j P^j(Y_i) \quad (1)$$

其中 π_j 为某一随机选择的个体属于第 j 组的概率,$P^j(Y_i)$ 表示该个体在第 j 组时得到 Y_i 的概率[8]。

GBTM 假定变量值在周期 T 内满足条件独立性,即对于给定的轨迹组 j 内的个体,在测量周期内 y_{it} 的值与之前各阶段相互独立,从而使得 $p^j(y_{it})$ 在其规范中不包括 y_{it} 的先验值,这有助于降低模型复杂性。因此:

$$P^j(Y_i)=\prod^{T}p^j(y_{it}) \quad (2)$$

其中 $p^j(y_{it})$ 表示组 j 中 y_{it} 的概率分布函数,其具体形式的选择取决于数据类型。若数据为量表数据,则使用删失正态分布(censored normal distribution)模型;若数据为一般正态分布,则将最大值和最小值设置在数据范围之外,同时采用删失正态分布模型;若数据为计数资料,则使用泊松分布(poisson distribution)模型;若数据为二分类数据,则使用基于逻辑分布(logit distribution)的模型。无论分布如何,其最终目的都是估计一组参数,使得 Y_i 的概率最大化,同时该参数具有定义轨迹形状和组成员概率的基本功能。

对于删失正态分布数据,时间和行为之间的联系是通过潜变量 y_{it}^* 来建立的,假设:

$$y_{it}^*=\beta_0^j+\beta_1^j Age_{it}+\beta_2^j Age_{it}^2+\beta_3^j Age_{it}^3+\varepsilon_{it} \quad (3)$$

其中,Age_{it} 表示在 t 时刻个体 i 的年龄,ε_{it} 表示服从均数为 0、标准差为 σ 的正态分布,β 用于确定轨迹的形状[7]。当测量值在最大值与最小值之间时,$y_{it}=y_{it}^*$。继而可得:

$$p^j(y_{it})=\frac{1}{\sigma}\Phi\left(\frac{y_{it}-\beta^j x_{it}}{\sigma}\right) \quad (4)$$

对于泊松分布数据,假设:

$$\mathrm{In}(\lambda_{it}^j)=\beta_0^j+\beta_1^j Age_{it}+\beta_2^j Age_{it}^2+\beta_3^j Age_{it}^3+\varepsilon_{it} \quad (5)$$

* 基金项目:天津市教委科研计划项目(2021KJ255);国家自然科学基金项目(81803333)
△通信作者:马骏,E-mail: junma@ tmu.edu.cn

其中，λ_{it}^{j}表示个体 i 在时间 t 内发生某事件（如犯罪）的期望次数，继而得到概率分布函数为：

$$p^{j}(y_{it})=\frac{\lambda_{jt}^{y_{it}}e^{-\lambda_{jit}}}{y_{it}!}(y_{it}=0,1,2,\cdots) \quad (6)$$

对于二分类数据，假设以组 j 的成员资格为条件，$p^{jt}(y_{it})$遵循二分类逻辑分布：

$$p^{jt}(y_{it})=\frac{e^{\beta j_0+\beta j_1 Age_{it}+\beta j_2 Age_{it}^2+\beta j_3 Age_{it}^3}}{1+e^{\beta j_0+\beta j_1 Age_{it}+\beta j_2 Age_{it}^2+\beta j_3 Age_{it}^3}} \quad (7)$$

组成员概率π_j，$j=1,2,3\cdots,J$，它是通过多项式 logit 函数估计的：

$$\pi_j=\frac{e^{\theta_j}}{\sum_{j=1}^{j}e^{\theta_j}} \quad (8)$$

其中θ_1归一化为零，用这种方式估计π_j可以确保其概率取值落在 0 和 1 之间。

2. 轨迹组数和多项式阶数的确定

确定最能代表发展轨迹异质性的群体数是 GBTM 的关键决策点之一。建模过程中，需要确定组数和多项式阶数，此过程需要分步进行，多次重复拟合，以确定最佳轨迹数目及形状。通常先固定多项式阶数以确定最佳组数，Nagin 建议此过程将所有组设置为 2 阶[9]，从 1 组开始逐渐增加组数，直至根据判断标准选择出最优组数；再固定组数以确定最优阶数，一般从最高阶开始，若高阶参数无统计学意义则去除，继续拟合低阶参数[7]。

通常选择贝叶斯信息准则（Bayesian information criterion，BIC）作为模型的判断标准之一，BIC 绝对值越接近 0，表明模型拟合效果越好[8]。此外还需要考虑其他因素：每组的平均后验概率>0.7；基于组成员后验概率的正确分类的优势（odds of correct classification，OCC）>5；模型在保证拟合效果的同时兼顾其简约性；组成员概率的置信区间不易过宽；同时每个轨迹组的样本数不宜过少等。

实例分析与 SAS 实现

1. PROC TRAJ 简介

PROC TRAJ 不属于 SAS 基本程序，可以通过网站进行下载：http://www.andrew.cmu.edu/user/bjones，其基本 SAS 程序如下。

```
PROC TRAJ DATA=ONE OUT=OF OUTPLOT=OP OUTSTAT=OS OUTEST=OE;
ID; VAR; INDEP;
MODEL; NGROUPS; ORDER; START;
MIN; MAX;
RORDER;  IORDER;
/* OUT：轨迹组的划分和组成员概率，OUTSTAT：宏使用的参数估计值，OUTPLOT：轨迹图数据，OUTEST：参数和协方差矩阵估计值；
ID：个体标识；VAR：个体在不同时间的测量值；INDEP：测量因变量时的时间或年龄；
MODEL：模型种类，CNORM 为删失正态分布数据使用的模型，ZIP 为零膨胀泊松分布使用的模型，LOGIT 为二分类逻辑分布使用的模型；
NGROUPS：拟合的轨迹数；ORDER：每个组的多项式（0=截距，1=线性，2=二次，3=立方）；START：参数起始值；
MIN/MAX：因变量的最小值和最大值（适用于 CNORM 模型）；
RORDER：CNORM 中每组的随机增长曲线参数（-1=无，0=截距，1=线性，2=二次，3=立方）；
IORDER：ZIP 中每组的多项式零膨胀概率 logit（0=截距，1=线性，2=二次，3=立方）；*/
```

2. 数据来源

本研究选取 B.Jones 网站（*http://www.andrew.cmu.edu/user/bjones）中提供的多项前瞻性研究数据，构建不同数据类型下的轨迹模型，进而找到因变量随时间变化的不同轨迹模式。使用模型之前，需要先将数据整理成表格格式，每一行数据代表一个受试者，包括 ID，VAR_{1-n}，T_{1-n}等变量，n 为最大随访次数，VAR 代表结局变量，T 代表随访时间或年龄，随访缺失的设为缺失值即可。

此外，由于模型的复杂性和可能存在的数据异常，一些模型可能难以拟合，因此确定起始值是必要的。若没有指定起始值，则该过程通过假设仅截距轨迹在因变量范围内均匀分布来提供默认起始值[10]。

3. 单轨迹模型

（1）当数据为删失正态分布，使用 CNORM 模型

本例选取蒙特利尔纵向研究的 916 名受试者的逆反行为数据，由教师在 6 岁和 10 至 15 岁时每年对其行为进行评估，打分范围为 0 到 10 分。使用 CNORM 模型进行轨迹的拟合。

表 1 最佳组数的模型迭代情况

组数	BIC 绝对值	最小组比例（%）
1	11657.91	100.00
2	10994.75	39.15
3	10867.23	23.78
4	10863.72	4.94

首先确定最佳组数，由表 1 可知，从 1 组增至 4 组，模型 BIC 绝对值依次减小，但 3 组增至 4 组时，最小组比例明显降低且 BIC 绝对值下降微弱，因此本研究的最佳轨迹组数为 3 组。确定最佳组数后，对各组最佳阶数进行拟合，程序和输出结果如下。

```
DATA OPPOSITN;
INPUT ID O1-O7; ARRAY T T1-T7(-0.4, 0, 0.1, 0.2, 0.3, 0.4, 0.5);
```

```
CARDS;
1    2    0    1    0    0    0    0
……
RUN;
PROC TRAJ DATA=OPPOSITN OUTPLOT=OP
OUTSTAT=OS OUT=OF OUTEST=OE ci95m;
ID ID; VAR O1-O7; INDEP T1-T7;
MODEL CNORM; MAX 10; ORDER 3 3 3;
RUN;
%TRAJPLOTNEW(OP, OS, '组别', '
', '逆反行为评分', '年龄')
/* %TRAJPLOTNEW:轨迹图及其 95%CI,"主
标题","副标题","纵坐标","横坐标" */
```

从每个轨迹组的立方次项开始拟合，结果如表 2。第一组中一次项参数是最高有效项（$t=-2.048$，$P=0.0406$），第二组中立方项参数为最高有效项（$t=2.900$，$P=0.0037$），第三组中二次项参数为最高有效项（$t=-6.754$，$P<0.0001$），因此多项式阶数分别为 1、3、2，即 ORDER 1 3 2。

表 2　CNORM 模型初次拟合结果

组别	参数	估计值	标准误	t 值	P 值
1	截距	−1.33549	0.27136	−4.922	<0.001
	线性	−2.73826	1.33706	−2.048	0.0406
	二次方	0.43320	1.74349	0.248	0.8038
	立方	1.26193	7.71462	0.164	0.8701
2	截距	2.50469	0.17199	14.563	<0.001
	线性	−3.78619	0.73103	−5.179	<0.001
	二次方	−7.52699	1.04028	−7.236	<0.001
	立方	12.17243	4.19736	2.900	0.0037
3	截距	5.63785	0.17768	31.731	<0.001
	线性	−1.24132	0.97142	−1.278	0.2014
	二次方	−9.15524	1.35554	−6.754	<0.001
	立方	9.65669	5.60303	1.723	0.0849

重新拟合得到最终的轨迹模型，结果如图 1 所示。28.6%的受试者表现出低水平的逆反行为（第 1 组）；47.9%的受试者表现出中等水平的对立行为（第 2 组）；23.5%的受试者初始具有较高水平对立行为且有上升趋势，升至 10 岁后逐渐下降（第 3 组）。模型的拟合优度为 BIC = −10873.39（$N=5834$），BIC = −10862.28（$N=916$）。

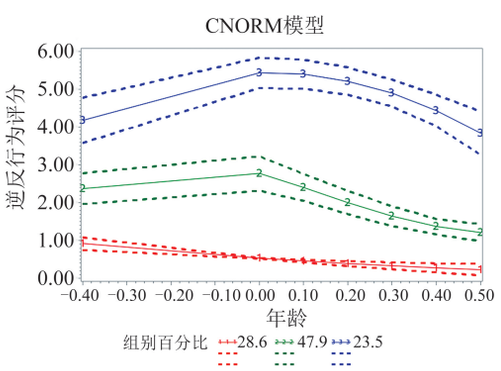

图 1　CNORM 模型最终拟合的 3 条轨迹及其 95%*CI*

（2）当数据为零膨胀泊松分布，使用 ZIP 模型

本例选取一项来自 195 名受试者的前瞻性研究数据，研究者每年从 8 岁到 32 岁的男性中收集其刑事犯罪记录的次数。按照前述方法确定最佳组数为 3 组，继而确定多项式阶数，程序和输出结果如下。

```
PROC TRAJ DATA=CAMBRDGE OUTPLOT=
OP OUTSTAT=OS OUT=OF OUTEST=OE ci95m;
ID ID; VAR C1-C23; INDEP T1-T23;
MODEL ZIP;  ORDER 3 3 3; IORDER 1;
/* IORDER 1 表明本例采用对所有组都通用的
线性函数构建额外的泊松零膨胀概率 */
RUN;
%TRAJPLOTNEW(OP, OS, 'ZIP 模型', ' ', '
犯罪次数', '年龄')
```

由表 3 可知，三组的多项式阶数均在立方次项有统计学意义（$t=2.827$，$P=0.0047$；$t=2.553$，$P=0.0107$；$t=3.372$，$P=0.0008$），因此不需要进行重复拟合。第 2 组受试者几乎从未被定罪，占比最大，为 71.0%；第 1 组受试者占比 21.8%，表现为低水平定罪率，仅青春期稍有增加；其余 7.2%受试者表现出高水平定罪率，并在青春期达到高峰，属于第 3 组。拟合优度为 BIC = −1013.90（$N=4485$），BIC = −988.82（$N=195$）。预测轨迹及其置信区间如图 2 所示。

表 3　ZIP 模型初次拟合结果

组别	参数	估计值	标准误	t 值	P 值
1	截距	−0.83628	0.21644	−3.864	0.0001
	线性	−0.44158	0.50057	−0.882	0.3777
	二次方	−1.22042	0.37282	−3.274	0.0011
	立方	1.72575	0.61042	2.827	0.0047
2	截距	−3.05466	0.35162	−8.687	<0.001
	线性	−1.91551	1.06973	−1.791	0.0734
	二次方	−2.89241	1.22845	−2.355	0.0186
	立方	4.61835	1.80906	2.553	0.0107
3	截距	0.53367	0.16098	3.315	0.0009
	线性	−1.35923	0.49387	−2.752	0.0059
	二次方	−1.88050	0.37792	−4.976	<0.001
	立方	2.21590	0.65713	3.372	0.0008

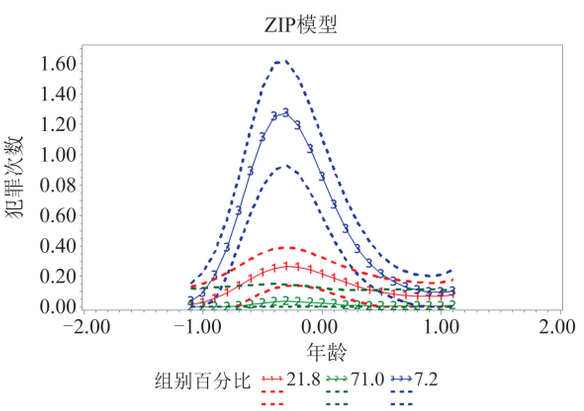

图 2　ZIP 模型最终拟合的 3 条轨迹及其 95%*CI*

4. 模型拓展

(1)添加协变量

向模型中添加协变量,可以识别与轨迹组成员相关的风险因素或保护因素,并探究其对轨迹组的影响。协变量分为普通协变量和时变协变量,在创建模型过程中,可分别使用“risk”和“tcov”语句进行拟合。纳入协变量通常使用三步法[9]:首先确定模型的最佳组数和多项式阶数;其次判断协变量与轨迹组的关联,可以采用 Wald 检验判断协变量是否有意义以及在不同轨迹组之间是否存在差异[1];最后纳入有意义的协变量,重新估计轨迹模型。

针对普通协变量,本例选用数据同样为 195 名受试者的前瞻性研究数据,首先对模型进行拟合;其次在模型中添加协变量,DARING 和 REARING 均为二分类变量,DARING 表示受试者在观察期间是否存在高冒险行为,REARING 表示受试者父母是否存在不良养育行为,1 表示存在。程序和输出结果如下。

```
PROC TRAJ DATA=CAMBRDGE OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;
ID ID; VAR C1-C23; INDEP T1-T23;
MODEL ZIP; NGROUPS 3; ORDER 3 3 3; IORDER 1;
RISK DARING REARING; REFGROUP 2;
RUN;
```

由表 4 可知,以第 2 组即低犯罪水平轨迹组作为参照,第 3 组与高冒险行为和不良养育行为的关系分别为 $t=2.795$, $P=0.0052$; $t=2.252$, $P=0.0244$,高冒险行为和不良养育行为均可以增加高犯罪群体的可能性,但不良养育行为对第 1 组的影响不具有统计学意义, $t=1.840$, $P=0.0658$。拟合优度为 BIC=−1019.87($N=4485$),BIC=−988.51($N=195$)。

表 4 添加普通协变量后模型拟合结果

组别	参数	估计值	标准误	*t* 值	*P* 值
1	常数	−1.79241	0.43056	−4.163	<0.001
	DARING01	1.03717	0.46220	2.244	0.0249
	REARING01	0.93878	0.51012	1.840	0.0658
2	基线	(0.00000)			
3	常数	−3.77198	0.68861	−5.478	<0.001
	DARING03	2.00459	0.71710	2.795	0.0052
	REARING03	1.47896	0.65667	2.252	0.0244

添加时变协变量可以判断某些临床干预(如药物治疗)是否会改变研究结果的发展过程[11]。本例中 F0~F10 为 T0~T10 时刻的时变协变量,假设已确定最佳组数和多项式阶数,程序和输出结果如下。

```
PROC TRAJ DATA=TRY OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE;
ID ID; VAR R0-R10; INDEP T0-T10;
MODEL LOGIT; ORDER 3 1 3 0;
TCOV F0-F10;
RUN;
%TRAJPLOT(OP, OS, 'LOGIT 模型', ' ', '缓解情况', '年龄')
```

由表 5 可知,协变量仅在第 1 组具有统计学意义($t=2.173$, $P=0.0299$),拟合优度:BIC=−922.14($N=1760$),BIC=−901.47($N=177$)。预测轨迹如图 3 所示。

表 5 添加时变协变量后模型拟合结果

组别	参数	估计值	标准误	*t* 值	*P* 值
1	截距	−3.67432	0.69974	−5.251	<0.001
	线性	−4.75473	2.57803	−1.844	0.0653
	二次方	4.25269	3.45594	1.231	0.2187
	立方	27.05611	13.25422	2.041	0.0414
	F001	0.27950	0.12864	2.173	0.0299
2	截距	0.59126	0.24709	2.393	0.0168
	线性	5.45379	0.65621	8.311	<0.001
	F002	0.02913	0.07923	0.368	0.7132
3	截距	5.44538	1.54251	3.530	0.0004
	线性	−6.45767	2.83188	−2.280	0.0227
	二次方	−19.31621	7.49360	−2.578	0.0100
	立方	34.39876	12.86777	2.673	0.0076
	F003	0.14847	0.23627	0.628	0.5298
4	截距	−0.28386	0.31727	−0.895	0.3711
	F004	0.06629	0.09418	0.704	0.4816

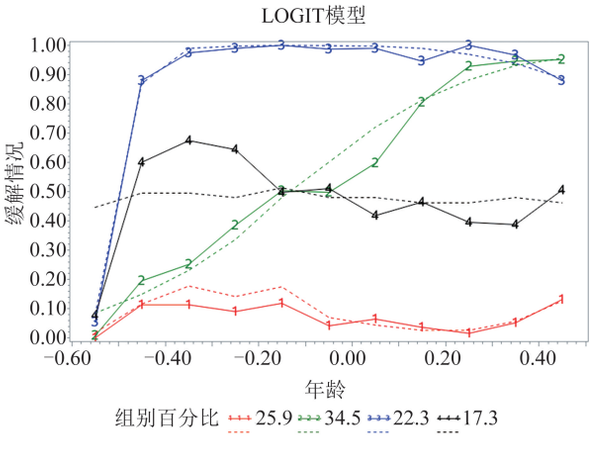

图 3 添加时变协变量后拟合的 4 个轨迹组

(2)双轨迹模型

双轨迹模型旨在识别总体异质性的前提下,以概率的形式分析两个指标间发展轨迹的关联性[12]。两个指标可以是时期发生变化,也可以不同时期变化。其步骤主要包括三步:①分别确定两个指标的单轨迹模型;②确定各轨迹组的成员概率;③求出两个指标各

轨迹组的两两关联概率。本例选用蒙特利尔纵向研究中的多动症和逆反行为量表拟合双轨迹模型。假设已经分别建立好两个单轨迹模型，输出结果见图 4。

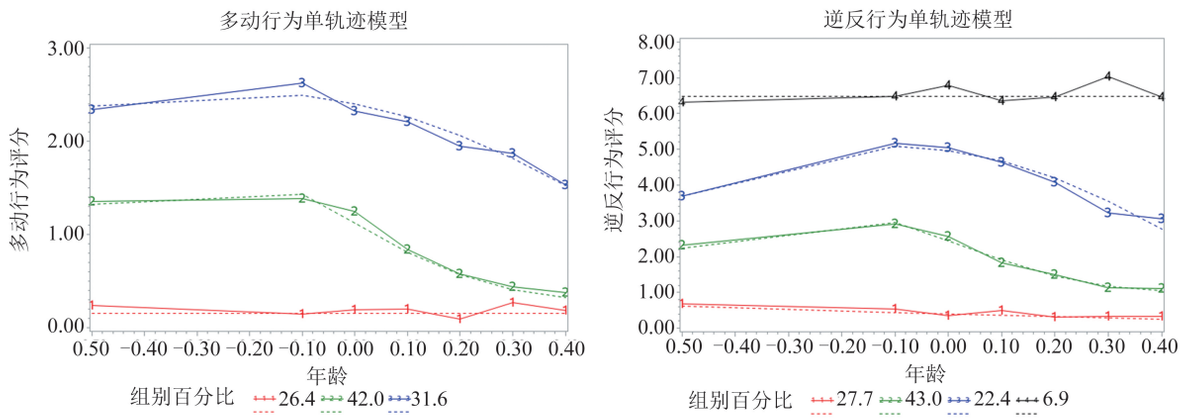

图 4　多动症及逆反行为的单轨迹模型

由图 4 可知，模型将多动症分为低水平（26.4%）、中等水平（42.0%）、高水平（31.6%）共 3 个轨迹组；将逆反行为分为低水平（27.7%）、中低水平（43.0%）、中高水平（22.4%）、高水平（6.9%）共 4 个轨迹组。进而计算两模型的联合概率，程序和输出结果如下。

```
PROC TRAJ DATA=ONE OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE OUTPLOT2=OP2 OUTSTAT2=OS2 ITDETAIL;
ID ID;
VAR H1-H7; INDEP T1-T7; MODEL CNORM; MAX 4; NGROUPS 3; ORDER 0 3 2;
VAR2 O1 - O7; INDEP2 T1 - T7; MODEL2 CNORM; MAX2 10; NGROUPS2 4; ORDER2 1 3 2 0;
RUN;
%TRAJPLOT(OP, OS, '多动行为单轨迹模型', ' ', '多动行为评分', '年龄')
%TRAJPLOT(OP2, OS2, '逆反行为轨迹模型', ' ', '逆反行为评分', '年龄')
```

表 6 为基于多动行为各组条件下的逆反行为各组概率，每列概率之和为 1。低水平多动行为组成为低水平逆反行为组的可能性为 99.9%；中等水平多动行为组成为中低水平逆反行为组的可能性为 96.9%；高水平多动行为组成为中高水平逆反行为组的可能性为 71.0%，成为高水平逆反行为组的可能性为 21.8%。

表 6　基于多动行为轨迹组条件下逆反行为轨迹组概率（%）

Group2（逆反行为）	Group1（多动行为）		
	低	中	高
低	99.9	3.1	0
中低	0	96.9	7.2
中高	0.1	0	71.0
高	0	0	21.8

表 7 为基于逆反行为条件下的多动行为各组的概率，每行概率合计为 1。低水平逆反行为组 95.4%归因于低水平多动行为组；中低水平逆反行为组 94.7%归因于中等水平多动行为组；中高水平逆反行为组 99.9%归因于高水平多动行为组；高水平逆反行为组 100%归因于高水平多动行为组。

表 7　基于逆反行为轨迹组条件下多动行为轨迹组概率（%）

Group2（逆反行为）	Group1（多动行为）		
	低	中	高
低	95.4	4.6	0
中低	0	94.7	5.3
中高	0.1	0	99.9
高	0	0	100

表 8 为多动行为和逆反行为的联合概率，总概率之和为 1。最终结果显示，两种行为的发展轨迹存在密切联系。多动行为低水平者成为逆反行为低水平者的可能性大于 99.9%，多动行为高水平者成为逆反行为中高或高水平者的可能性共计 92.8%。研究对象中有 26.4%同时属于多动行为和逆反行为低水平组，40.7%同时属于多动行为和逆反行为中等（中低）水平组，29.3%同时属于多动行为和逆反行为高（中高）水平组。

表 8　逆反行为和多动行为轨迹组联合概率（%）

Group2（逆反行为）	Group1（多动行为）		
	低	中	高
低	26.4	1.3	0
中低	0	40.7	2.3
中高	0	0	22.4
高	0	0	6.9

讨　论

基于分组的轨迹模型通过区分不同轨迹的个体群组，在充分利用纵向数据的基础上，将结果以易于解释

的形式展现,为解决纵向数据复杂性和结果的可解释性之间的平衡问题提供了一个有价值的工具[13]。本研究系统阐述了不同数据类型的模型基本原理及其SAS软件实现;并在此基础上将协变量纳入模型,观察影响行为轨迹的因素;同时,本研究还介绍了双轨迹模型,以概率形式表达两个纵向指标之间的关联。

GBTM相较于经典的发展轨迹分析有其独特的优势,但在实际应用中应当注意一系列问题。首先,轨迹组数和多项式阶数的确定需要结合专业知识谨慎选择,兼顾其实用性和可解释性[14]。此外,GBTM对于缺失数据的处理具有一定的稳健性,但模型拟合时应考虑到缺失值的类型和比例[11]。同时,由于GBTM假设不同组间的固定效应不同,但组内具有相同的固定效应,即认为同组内个体不存在异质性。而在实际应用中,这一假设往往会造成一定的信息损失。潜在类别混合模型(latent class mixed model, LCMM)[15]是GBTM的更一般形式,他兼顾组间的差异性和组内个体的随机效应,从而达到了对轨迹模型的进一步优化。然而无论何种轨迹分析方法,继续加强模型参数选择的规范及评价标准的统一都尤为重要。研究者也应当在应用时充分了解不同方法的原理及适用条件,避免主观因素,提高研究结论的可靠性。

参 考 文 献

[1] Jones BL, Nagin DS. Advances in Group-Based Trajectory Modeling and an SAS Procedure for Estimating Them. Sociological Methods & Research, 2007, 35(4): 542-571.

[2] Kakoly NS, Earnest A, Moran LJ, et al. Group-based developmental BMI trajectories, polycystic ovary syndrome, and gestational diabetes: a community-based longitudinal study. BMC Medicine, 2017, 15(1): 195.

[3] Dekker MC, Ferdinand RF, van Lang ND, et al. Developmental trajectories of depressive symptoms from early childhood to late adolescence: gender differences and adult outcome. Journal of Child Psychology and Psychiatry, and Allied Disciplines, 2007, 48(7): 657-666.

[4] Huber KA, Frazier PA, Alper HE, et al. Trajectories of posttraumatic stress symptoms following the September 11, 2001, terrorist attacks: A comparison of two modeling approaches. Journal of Traumatic Stress, 2022, 35(2): 508-520.

[5] Alhazami M, Pontinha VM, Patterson JA, et al. Medication Adherence Trajectories: A Systematic Literature Review. Journal of Managed Care & Specialty Pharmacy, 2020, 26(9): 1138-1152.

[6] Linden A. Using group-based trajectory modelling to enhance causal inference in interrupted time series analysis. Journal of Evaluation in Clinical Practice, 2018, 24(3): 502-507.

[7] 张晨旭, 谢峰, 林振, 等. 基于组轨迹模型及其研究进展. 中国卫生统计, 2020, 37(6): 946-949.

[8] 冯国双, 于石成, 胡跃华. 轨迹分析模型在流行病学研究中的应用. 中华流行病学杂志, 2014, 35(7): 865-867.

[9] Nagin D. Group-Based Modeling of Development. Harvard University Press, 2009.

[10] Jones BL, Nagin DS, Rorder K. A SAS Procedure Based on Mixture Models for Estimating Developmental Trajectories. Sociological Methods & Research, 2001, 29(3): 374-393.

[11] Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. Annual Review of Clinical Psychology, 2010, 6: 109-138.

[12] 郭剑, 高洪艳, 王媛. 基于组基础模型的HIV/AIDS发病和死亡纵向资料双轨迹分析. 中国卫生统计, 2021, 38(3): 358-362.

[13] Nagin DS. Group-based trajectory modeling: an overview. Ann Nutr Metab, 2014, 65(2-3): 205-210.

[14] Nagin DS, Jones BL, Passos VL, et al. Group-based multi-trajectory modeling. Stat Methods Med Res, 2018, 27(7): 2015-2023.

[15] 殷畅, 武振宇, 郑雪莹. 潜在类别混合模型及其在纵向数据轨迹分析中的应用. 中国卫生统计, 2022, 39(4): 538-541.

(责任编辑:邓 妍)