

# 轨迹分析模型在流行病学研究中的应用

冯国双 于石成 胡跃华

**【导读】** 通过对60名青少年4次脂肪发育数据的分析,介绍轨迹分析模型在流行病学研究中的应用。结果显示,正常脂肪组、偏高脂肪组、超高脂肪组具有不同的发展趋势。轨迹分析模型在拟合异质性追踪数据方面具有较好的优势。

**【关键词】** 轨迹分析模型;潜分类增长模型

**The application of trajectory model in epidemiological research** Feng Guoshuang, Yu Shicheng, Hu Yuehua. National Center for Public Health Surveillance and Information Services, Chinese Center for Disease Control and Prevention, Beijing 102206, China

Corresponding author: Feng Guoshuang, Email: glxfgh@163.com

This work was supported by a grant from the Project of Comprehensive Prevention and Treatment of Cancer in Huaihe River Basin from the Special Fund of the Central Government (No. 1310800003).

**【Introduction】** This article introduced the application of trajectory model in epidemiological study on fat development in sixty teenagers. Results showed that fat developing trend on three trajectories-normal fat group, high fat group and very high fat group, with independently different developing trend. Trajectory model seemed a suitable model for fitting the longitudinal data with heterogeneity.

**【Key words】** Trajectory model; Latent class growth model

流行病学研究常关注某现象随时间变化的特征,如青少年的生长发育特征、患者病情变化等,这类数据通常称为追踪数据。对于这类资料,常用方法是重复测量方差分析、生长曲线模型、多水平模型等。尽管这些方法可以有效地对总体趋势进行描述,但存在一定局限,即假定总体发展趋势相同。而这一假定在现实中往往不能满足,许多现象随时间变化,趋势具有异质性,如对于青少年体格发育,一部分人群发育很快,而另一部分人发育较慢,有的人一开始较慢而后期变快。对此采用上述方法就不能很好地描述这种趋势及差异。

为解决这一问题,不少学者开发了一些新方法,目前较为常用的是轨迹分析模型(trajectory model)<sup>[1,2]</sup>,该法可以针对不同发展趋势的亚组进行描述和拟合。本文对轨迹分析模型进行简要介绍,并结合实例说明该方法在流行病学研究中的具体应用。

## 基本原理

假定  $Y_i = \{y_{i1}, y_{i2}, y_{i3}, \dots, y_{iT}\}$  表示第  $i$  个体在第  $t$

时间点上的一个纵向观测序列值,  $P(y_i)$  表示  $y_i$  的概率。由于轨迹分析模型假定总体中存在  $j$  个潜在的发展趋势不同的亚组,因此

$$P(y_i) = \sum_j \pi_j P^j(Y_i)$$

其中,  $\pi_j$  表示第  $j$  组的概率,  $P^j(Y_i)$  表示在假定组成员(membership)属于  $j$  组这一条件下  $Y_i$  的概率。

该模型同时假定在确定了组成员属于  $j$  组的条件下,随机变量  $y_{it} (t=1, 2, \dots, T)$  是相互独立的。因此

$$P^j(Y_i) = \prod_{t=1}^T p^j(y_{it})$$

应根据数据的类型确定  $p^j(y_{it})$  的形式。本研究变量为脂肪量,为具有一定范围的定量资料,因此可定义为删截正态分布(censored normal distribution)。

总体上,轨迹分析模型的主要目的是探索总体中包含有多少发展趋势不同的亚组,并确定各亚组的发展轨迹。该模型通过估计在给定亚组  $j$  组成员条件下  $y_{it}$  的条件概率分布函数,确定个体发展轨迹。

轨迹分析模型的关键是确定恰当的亚组数目以及各亚组的形状,通常可根据以下几个指标来确定<sup>[3]</sup>:①贝叶斯信息标准(BIC),BIC值越接近0,表示模型拟合越好。②贝叶斯因子对数值,该值近似

DOI: 10.3760/cma.j.issn.0254-6450.2014.07.024

基金项目:中央财政专项淮河流域癌症综合防治项目(1310800003)

作者单位:102206 北京,中国疾病预防控制中心公共卫生监测与信息服务中心

通信作者:冯国双, Email: glxfgh@163.com

等于2个比较模型BIC差值的2倍。一般该值 $>6$ 表示有很强的证据,可以接受复杂模型;该值 $<2$ 时,表明不足以接受复杂模型,应采用简单模型。③平均验后分组概率(AvePP),该指标反映了根据轨迹分组后的亚组内成员与该轨迹的符合程度,通常 $>0.7$ 为可接受的标准。

轨迹分析模型目前可通过SAS软件的proc traj过程来实现,该过程是由Jones等<sup>[4]</sup>开发的宏程序,专门用于分析异质性发展轨迹。

### 实例分析

本研究数据来自于一项针对发育期女生脂肪量的长期随访研究,该研究对一群10岁的女生测量其脂肪量,并分别于2、4、6年后分别再次测量,以观察脂肪量变化情况。本例从中随机选出60例4次观测均完整的数据作为本次实例分析用。

首先从研究对象中随机抽取了6例受试者的数据,绘制其脂肪量随时间变化的趋势。不同研究对象的发展趋势各有不同(图1),提示可能存在异质性。因此可以考虑采用轨迹分析模型寻找不同发展趋势的亚组。

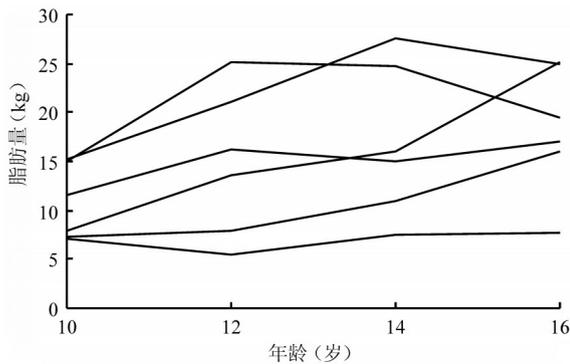


图1 6名观测对象脂肪量随时间变化的趋势

考虑到本研究数据较少,并结合实际情况,分别对1、2、3组的模型进行拟合( $\geq 4$ 组情况未探索)。其BIC值分别为-820.30、-735.62、-707.09,3组的BIC值比单组、2组升高的值远远 $>6$ ,可以较明显地看出分3组拟合效果最佳。而且从AvePP来看,3组的验后概率分别为0.98、0.99、1.00,提示分3组效果较好。

表1给出了3个亚组各参数的估计结果,图2给出了3个亚组的发展轨迹及各组的比例。各亚组发展轨迹的拟合均从高阶(3阶)开始,高阶不显著时,则继续拟合低阶项。拟合结果显示,第1组的3阶和2阶均无统计学意义,直线趋势有统计学意义。第2

组和第3组的3阶均无统计学意义,而2阶和1阶均有统计学意义。

表1 3个亚组各参数的估计结果

亚组	参数	参数估计	$s_e$	$t$ 值	$P$ 值
1	intercept	7.184 87	0.431 42	16.654	0.000 0
	linear	0.099 98	0.008 89	11.252	0.000 0
	quadratic	0.000 00	0.000 00	0.000	1.000 0
2	intercept	16.577 69	0.945 30	17.537	0.000 0
	linear	0.296 37	0.054 72	5.416	0.000 0
	quadratic	-0.002 30	0.000 60	-3.815	0.000 2
3	intercept	28.077 96	2.489 22	11.280	0.000 0
	linear	0.406 75	0.144 63	2.812	0.005 3
	quadratic	-0.003 29	0.001 59	-2.071	0.039 5

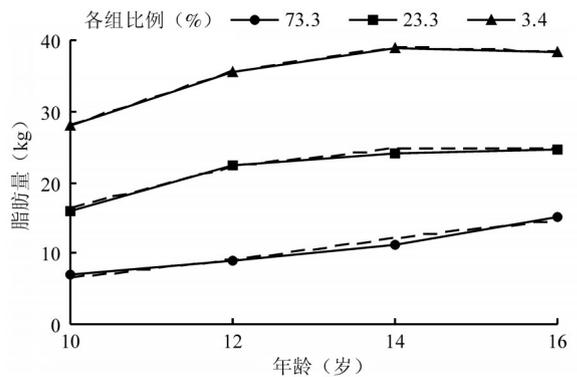


图2 3个亚组的脂肪量发展轨迹图

第1组人群占了绝大多数(73.3%),反映了正常女生的脂肪增长情况,称为正常脂肪组。这一亚组初始脂肪量最低(intercept=7.18487),但一直呈直线上升趋势(slope=0.09998)。

第2组人群占23.3%,这一人群呈二次曲线发展趋势,称为偏高脂肪组。初期脂肪量较高(intercept=16.57769),后2年呈直线上升趋势(linear slope=0.29637),但2年后增长速度下降(quadratic slope=-0.00230)。

第3组人群比例最低,仅为3.3%,这部分人虽然少,但代表了某些发育较为特殊的女生,称为超高脂肪组。这一亚组初期脂肪量便很高(intercept=28.077 96),而且上升速度较快(linear slope=0.406 75),但到4年后增长速度开始下降(quadratic slope=-0.00329)。

### 讨 论

轨迹分析模型在有的领域也称为潜分类增长模型(LCGM),其原理就是假定总体存在异质性,可分为多个潜在的亚组或类别,目的是探索各亚组的发展轨迹,并结合专业知识分别赋予各亚组较为实际的含义名称。

轨迹模型的分析是一个动态过程,要得到一个最佳模型,不仅需要确定亚组数目,还需要确定各亚组的发展轨迹形状。通常情况下,亚组的确定需同时考虑统计学评价指标和专业知识。如本分析案例中,不管是从统计学还是专业来看,单组都不是最佳选择。如果分 4 组,结果主要是把第 1 组人群又分两类(结果未列出),形状完全一致,只是截距的不同,从专业上来看,并无太多实际意义。因此本例最终选择分为 3 种发展轨迹。

各亚组形状确定通常先从高阶开始,如果高阶参数统计学意义不显著,则去掉高阶,继续拟合低阶参数。为了保证模型的简洁性,无统计学意义的高阶参数应从模型中剔除,但线性参数不管其统计学意义是否显著,一般都应包含在模型中<sup>[3]</sup>。

轨迹分析模型可以较好地拟合异质性总体,该模型假定总体内存在不同发展轨迹的亚组,但在同一亚组内假定是同质的,即假定同一亚组内的个体具有相同的截距和斜率。这一假定有时在实际中难以满足,针对这一问题,Muthén 和 Muthén 等<sup>[5]</sup>提出了增长混合模型(GMM),该模型可以看作是轨迹分析模型的扩展,它允许同一亚组内个体的增长参数(截距、斜率等)存在变异。由于增长混合模型允许同一亚组内存在个体变异,因此其拟合指标往往优于轨迹分析模型,但由此也带来估计的复杂。实际应用中,通常可先对数据采用轨迹分析模型确定亚组及轨迹形态,然后可观察各亚组内的个体是否相

对同质,如果形态大致相同,则采用轨迹分析模型即可。如果发现同一亚组内个体异质性较大,可进一步采用增长混合模型进行拟合。

轨迹分析模型在国外不少领域已取得较好应用,尽管有学者提出了该模型在应用中的一些局限性<sup>[6]</sup>,但该模型仍是一个积极发展的领域,而且不断有新的研究成果出现<sup>[7]</sup>。目前轨迹分析模型已可以通过 SAS 软件的 proc traj 过程及 M-Plus 软件来实现,相信在今后会有更广泛的推广和应用空间。

#### 参 考 文 献

- [1] Nagin D. Analyzing developmental trajectories: a semiparametric, group-based approach [J]. *Psychol Methods*, 1999, 4(2): 139-157.
- [2] Nagin D. *Group-based Modeling of Development* [M]. MA: Harvard University Press, 2005.
- [3] Andruff H, Carraro N, Thompson A, et al. Latent class growth modelling: a tutorial [J]. *Tutor Quantit Methods Psychol*, 2009, 5(1): 11-24.
- [4] Jones BL, Nagin DS, Roeder K. A SAS procedure based on mixture models for estimating developmental trajectories [J]. *Sociol Method Res*, 2001, 29(3): 374-393.
- [5] Muthén B, Muthén LK. Integrating person-centered and variable-centered analyses: growth mixture modeling with latent trajectory classes [J]. *Clin Exp Res*, 2000, 24(6): 882-891.
- [6] Bauer DJ, Curran RJ. Distributional assumptions of growth mixture modes: implications for overextraction of latent trajectory classes [J]. *Psychol Methods*, 2003, 8(3): 338-363.
- [7] Jones BL, Nagin DS. Advances in group-based trajectory modeling and an SAS procedure for estimating them [J]. *Sociol Methods Res*, 2007, 35(4): 542-571.

(收稿日期: 2014-01-07)

(本文编辑: 万玉立)