

· 研究原著 ·

文章编号: 1000-2790(2000)06-0676-03

## 拟合生长标准百分位数曲线的不同方法比较

尚磊<sup>1</sup>, 徐勇勇<sup>1</sup>, 侯茹兰<sup>2</sup>, 张水平<sup>3</sup>, 周引荣<sup>3</sup>, 陈长生<sup>1</sup>( <sup>1</sup> 第四军医大学预防医学系卫生统计学教研室, 陕西 西安 710033, <sup>2</sup> 西安医科大学儿少卫生教研室, <sup>3</sup> 西安市妇幼保健院)

关键词: 青春期医学; 发育生物学; 模型; 统计学

中图分类号: R179 文献标识码: A

**摘要:** 目的 寻找适合于拟合各种人体测量指标的生长标准曲线的方法和途径。方法 采用损失函数(loss function)和残差平方和(residual sum of square)对加权三次样条法(weighted cubic smoothing spline)和 $\lambda$ -中位数-变异系数法( $\lambda$ -median-coefficient of variation, LMS)进行对比研究。结果 计算了两种方法拟合各组的身高(height)、体质量(body mass)、体块指数(body mass index, BMI)的损失函数值和残差平方和, 给出了两种方法拟合的百分位数曲线。结论 LMS法简单易实现, 拟合的百分位数曲线光滑, 拟合误差小。LMS法比加权三次样条更适合于拟合身高、体质量、体块指数百分位数曲线。

## Comparison of different approaches of fitting centile curves for growth standard

SHANG Lei<sup>1</sup>, XU Yong-Yong<sup>1</sup>, HOU Ru-Lan<sup>2</sup>, ZHANG Shui-Ping<sup>3</sup>, ZHOU Yin-Rong<sup>3</sup>, CHEN Chang-Sheng<sup>1</sup>

<sup>1</sup>Department of Health Statistics, Faculty of Preventive Medicine, Fourth Military Medical University, Xi'an 710033, China, <sup>2</sup>Department of Health for Children and Adolescence, Xi'an Medical University, <sup>3</sup>Xi'an Maternal and Child Health Station

**Keywords:** adolescent medicine; developmental biology; models; statistics

**Abstract:** **AIM** To find out a method to construct growth curves for various anthropology index. **METHODS** Loss function and residual sum of square were used to compare the

fitting results of LMS method with weighted cubic smoothing spline. **RESULTS** Loss function and residual sum of square for each method when applied to constructing body mass, height and BMI centile curves for children aged 0-18 years in Xi'an were calculated. Centile curves fitted by LMS method and weighted cubic smoothing spline were presented. **CONCLUSION** LMS is a simple method. Curves fitted by it are much smoother and the fitting errors are much fewer. Therefore LMS is more suitable for fitting centile curves for body mass, height and BMI than weighted cubic smoothing spline.

## 0 引言

年龄别百分位数广泛用于儿科临床和儿保工作中, 用以监测儿童生长发育及营养状况。建立年龄别百分位数的方法很多, 大体可分为两类: 一类是对资料的总体分布不作要求, 计算实际百分位数, 然后进行修匀; 另一类百分位数的计算是假定资料服从正态分布或经一个合适的转换后可成正态分布<sup>[1-3]</sup>。本文用代表两类方法的加权三次样条和LMS法分别拟合西安市0~18岁儿童青少年身高、体质量和体块指数百分位数曲线, 以比较两种方法在儿童青少年体格发育指标百分位数曲线拟合中的适用性。

## 1 资料和方法

**1.1 资料** 取自1995年全国0~6岁儿童营养与发育调查和1995年全国学生体质与健康调研西安数据。将0, 1, 2, 3, 4, 5, 6, 8, 10, 12, 15, 18, 21, 24月与2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18岁作为年龄中位数。0~6岁每组200人, 7岁以后每组100人, 城乡、男女一致。

**1.2 加权三次样条方法** 设某区间 $[a, b]$ 上有实数 $t_1, t_2, \dots, t_n$ 且满足 $a < t_1 < t_2 < \dots < t_n < b$ ,  $f(x)$ 为定义在 $[a, b]$ 上的函数, 若 $f(x)$ 满足以下两个条件: ①在 $[a, t_1], [t_1, t_2], \dots, [t_n, b]$ 每一区间上 $f(x)$ 为三次多项式。②函数 $f(x)$ 及其前二阶导数在 $t_i (i=1, 2, \dots, n)$ 处都连续, 点 $t_i$ 称为样条函数的节点。三次样条函

收稿日期: 1999-11-10; 修回日期: 1999-12-27

基金项目: 全军医药卫生科研基金课题(96M0987)

作者简介: 尚磊(1968-), 男(汉族), 陕西省长武县人。硕士, 讲师。

Tel: (029) 3374853 Email: xuyongy@fmmu.edu.cn

数的表达形式有多种,不同形式间等价,令  $t_0 = a, t_{n+1} = b$ , 则称

$$f(x) = d_i(x - t_i)^3 + c_i(x - t_i)^2 + b_i(x - t_i) + a_i$$

$t_i$  $x$  $t_{i+1}$

为三次样条函数. 假定  $t_i$  时残差的权重赋为  $w_i, w_i > 0$ , 则加权残差平方和为  $\sum w_i (y_i - f(t_i))^2$ , 上述选择函数的惩罚平方和为

$$S(f) = \sum w_i (y_i - f(x_i))^2 + \lambda \int_a^b (f''(x))^2 dx,$$

对于给定的光滑参数  $\lambda (\lambda > 0)$  则使  $S(f)$  最小的估计函数  $f(x)$  称为惩罚最小二乘估计. 光滑参数  $\lambda$  可由  $\lambda = CQ^3/1000$  算出,  $C$  为给定的常数,  $Q$  为解释变量的四分位数间距.

**1.3 LMS 方法** 采用 LMS 法拟合百分位数曲线. LMS 法是用中位数( $M$ )、变异系数( $S$ ) 及将资料转换成正态所需要的度量偏度的  $Box-Cox$  转换的幂( $L$ ) 概括了变量在每个年龄的分布. 这 3 个数值随年龄变化, 特别是  $M$  随年龄变化的曲线就是变量的第 50 个百分位数曲线. 其他百分位数可用下式获得

$$C_{100\alpha}(t) = M(t) [1 + L(t)S(t)Z_{\alpha}]^{1/L(t)} \tag{1}$$

其中  $Z_{\alpha}$  为尾部面积  $\alpha$  的正态离差,  $C_{100\alpha}$  为  $Z_{\alpha}$  所对应的变量的百分位数(如第 25 个百分位数  $\alpha = 0.25, Z_{\alpha} = -0.67$ ),  $t$  为年龄,  $L(t), S(t), M(t)$  和  $C_{100\alpha}(t)$  为年龄为  $t$  时所对应的每条曲线的相应值.

**1.4 损失函数** 是百分位数的期望比例( $P_{97}, P_{90}$ ) 与实际比例(观察值中低于百分位数的拟合值的比例,

如  $P_{97}$  的拟合值与该年龄组所有观察值比较, 观察值中低于拟合值的观察值占的比例) 的差, 表示拟合后分布与实际分布的差别, 损失函数越小, 拟合后分布与实际分布越接近, 损失函数的计算公式如下:

$$Loss\ function = \{ \sum \sum (p_{ia} - \pi_i)^2 / [60\pi_i(1 - \pi_i)] \} \tag{2}$$

其中:  $\pi_i$  是第  $i$  个百分位数的期望比例, 即  $\pi_3 = 0.03, \pi_5 = 0.05, \pi_{10} = 0.10$  等.  $p_{ia}$  是第  $a$  个年龄组中低于其第  $i$  个百分位数拟合值的观察比例.

**1.5 残差平方和** 是拟合值与实际值之间残差的平方和, 残差平方和越小, 拟合值与实际值越接近, 拟合结果越准确. 计算公式为:

$$residue\ sum\ of\ square = \sum \sum (\hat{Y}_{ia} - \hat{y}_{ia})^2 \tag{3}$$

其中,  $Y_{ia}, \hat{y}_{ia}$  分别为第  $a$  个年龄组第  $i$  个百分位数的拟合值和实际值.

**1.6 计算方法** 采用 NoSA 统计软件计算年龄别百分位数, 用 SAS 软件拟合加权三次样条; 用 Cole 教授提供的 LMS 法计算程序完成 LMS 法拟合过程, 用 EXECL 软件绘图.

2 结果

不同方法拟合的损失函数和残差平方和见 Tab 1. 损失函数值表示拟合后分布与实际分布的差别, 残差平方和表示拟合的误差. 因此二者可用于评价拟合方法的准确性. 从 Tab 1 可见, 身高、体质量、体块指数的拟合效果, 各组均以 LMS 法的损失函数和残差平方和最小.

表 1 不同方法拟和身高、体质量、体块指数的损失函数和残差平方和

Tab 1 Loss function and residual sum of square of two methods for fitting body mass, height and BMI centile curves

Group		LMS method		Weighted cubic smoothing spline	
		Loss function	Residual sum of square	Loss function	Residual sum of square
Height	Urban boys	0.0091	17.997	0.0100	22.440
	Urban girls	0.0086	20.547	0.0093	23.051
	Rural boys	0.0076	15.647	0.0082	22.394
	Rural girls	0.0089	17.894	0.0100	19.151
Body mass	Urban boys	0.0070	9.800	0.0109	28.861
	Urban girls	0.0049	9.173	0.0098	20.597
	Rural boys	0.0075	8.800	0.0101	17.494
	Rural girls	0.0065	12.755	0.0083	16.130
BMI	Urban boys	0.0079	1.677	0.0123	7.692
	Urban girls	0.0071	1.655	0.0092	6.347
	Rural boys	0.0081	1.396	0.0086	5.595
	Rural girls	0.0064	1.493	0.0087	5.782

3 讨论

LMS 法能保证通过 *Box-Cox* 转化使资料达到渐近正态, *L* 曲线是该法的核心, 利用这种在偏度上的信息, 使得估计极端的百分位数比简单的“排序和计数”法更好, 每个百分位数估计的标准误很小. *M* 曲线恰好是中位数曲线, *S* 曲线代表变异系数, *L* 曲线和 *M*、*S* 曲线提供了用个体测量值计算精确百分位数和产生标准离差的所有必须信息. *L*、*M*、*S* 3 个参数决定曲线形状, 而这 3 条曲线遵循儿童期各自的变化规律, 因此, 曲线上的任一突起或凹下都是资料真实特征的反映. LMS 法拟合的曲线光滑, 与资料接近, 能保证邻近百分位数间的距离. 不好之处是计算的过程复杂, 需要专门程序实现<sup>[4, 5]</sup>.

三次样条法的缺点是节点数不易估计, 在用于 NCHS(national center for health statistics) 百分位数时结果不满意, 由于百分位数的标准误在分布的尾部增加很快, 因此经验百分位数越极端拟合结果越不准确, 三次样条适合于经验百分位数相对光滑时. 用加权样条拟合百分位数曲线, 虽然考虑了资料的变异的方差, 但拟合结果仍不够理想<sup>[6- 8]</sup>.

从损失函数和残差平方和的分析可见, 加权样条法拟合的身高、体质量、BMI 的损失函数和残差平方和均比 LMS 法拟合的大. 从曲线形态和光滑度可见, LMS 法拟合的 BMI 曲线与其他国家和地区的 BMI 曲线形态一致<sup>[9]</sup>, 且更符合儿童青少年体形发育规律, 如城女组在青春期末 *P*<sub>97</sub> 百分位数出现下降趋势, *P*<sub>90</sub> 平稳变化, 这与城市偏胖女生有主观减肥现象的情况一致, 而加权三次样条拟合的各条曲线在青春期末仍处于上升趋势, 这与实际不符<sup>[8]</sup>. 而在身

高、体质量百分位数曲线的拟合中, 加权三次样条法拟合的边缘百分位数存在轻微抖动, 不够光滑<sup>[7, 8, 10]</sup>. 曲线拟合的要求是: 曲线光滑; 与经验百分位数接近, 因此, 相比而言, LMS 更适合于拟合身高、体质量、BMI 百分位数曲线.

参考文献:

[1] Bonellie SR, Raab GM. A comparison of different approaches for fitting centile curves to birthweight data[J]. *Stat Med*, 1996; 15: 2657- 2667.

[2] Gasser T, Molinari L, Roos M. Methodology for the establishment of growth standards[J]. *Horm Res*, 1996; 45(Suppl 2): 2- 7.

[3] Healy MJR, Rasbash J, Yang M. Distribution-free estimation of age-related centiles[J]. *Ann Hum Biol*, 1988; 15(1): 17- 22.

[4] Cole TJ. Fitting smoothing centile curves to reference data[J]. *J R Stat Soc A*, 1988; 151: 385- 415.

[5] Cole TJ, Green PJ. Smoothing reference centile curves: The LMS method and penalized likelihood[J]. *Stat Med*, 1992; 11: 1305- 1319.

[6] 陈长生. 非参数回归和生长曲线统计分析方法研究及其医学应用[D]. 西安: 第四军医大学, 1998: 35- 45.

[7] 陈长生, 徐勇勇, 尚 磊. 儿童生长曲线的光滑样条和核估计拟合[J]. 中国卫生统计, 1997; 14(4): 1- 3.

[8] 尚 磊. 生长标准图 LMS 平滑方法的研究与应用[D]. 西安: 第四军医大学, 1999: 47- 74.

[9] Cole TJ, Freeman JV, Preece MA. Body mass index reference curves for the UK, 1990[J]. *Arch Dis Child*, 1995; 73(5): 25- 29.

[10] 尚 磊, 徐勇勇, 陈长生 *et al.* 拟合体质量百分位数的加权三次样条[J]. 第四军医大学学报, 1999; 20(8): 684- 685.

编辑 何扬举

· 文 摘 · Bax 基因对人胃癌耐药细胞多药耐药性的逆转作用

[ 赵燕秋, 肖 冰, 樊代明. 中华消化杂志, 2000; 20(1): 8- 10]

目的 研究凋亡相关基因 Bax 对胃癌多药耐药性的逆转作用. 方法 构建含有人 Bax cDNA 全长的真核表达载体 pBK-Bax. 经脂质体导入缺乏 Bax 蛋白表达的人胃癌多药耐药细胞系 SGC7901/VCR 中. Western 印迹观察 Bax 基因在转导细胞中的表达, MTT 法检测 Bax 转导细胞与空载体转导细胞对化疗药物的敏感性. 结果 成功构建了 Bax 的真核表达载体, Bax 转导细胞能稳定表达 Bax 蛋白, 与空载体转导细胞相比, Bax 转导细胞对长春新碱(*P* < 0.01). 结论 Bax 基因转导对胃癌多药耐药细胞 SGC7901/VCR 的多药耐药性具有逆转作用.

(李之源)