[文章编号] 1003－4684(2026)04-0030-04

基于 MERS 模型的高校学生超重和肥胖触发年龄的统计分析

郑江松[1]，吴士宝[2]，罗幼喜[2]

(1 湖北工业大学电气与电子工程学院，湖北 武汉 430068；2 湖北工业大学理学院，湖北 武汉 430068)

[摘 要] BMI(体质指数)是衡量身体超重和肥胖的重要指标，青少年 BMI 随年龄演变的规律是对青少年进行肥胖干预的重要依据。研究高校学生 BMI 随年龄变化而变化的特征，构建出高校学生 BMI-年龄曲线，在此基础上，对高校学生超重和肥胖触发年龄进行统计分析。针对已有“分位数回归方法”的不足，根据回归样条线混合效应模型(MERS 模型)，利用 AIC 指标选择合适的结点数和平滑参数，通过数据分析得出中学和大学是预防学生超重和肥胖的关键时期。

[关键词] BMI；MERS 模型；超重；肥胖；触发年龄

[中图分类号] TP391；F830.589 [文献标识码] A

1 概述

目前，全球范围内高校学生超重和肥胖率急剧增长，导致学生的身体健康状况不容乐观，这一现象已引起了世界各国学者的高度关注[1-5]。针对高校学生出现的超重和肥胖现象，世界各国都积极采取各种措施进行控制，但效果不佳[6-7]。判断超重和肥胖的标准有很多，例如有腰围—臀围指数、体质指数、腰围—身高指数等，但世界卫生组织推荐的指标主要是体质指数，体质指数又称体重指数，也称身体质量指数，简称 BMI(body mass index).该指标是通过体重(kg)与身高(m)平方的比值计算得来，即

$$\mathrm{BMI}=\frac{\text{体重}(kg)}{[\text{身高}(m)]^2} \tag{1}$$

式(1)所计算出的结果在一定程度上可以反映人体密度，BMI 具有测量简单，计算方便的特点，现在被普遍用于评价人体的营养状况、胖瘦程度或身体发育水平，是国际上常用的衡量人体胖瘦程度以及是否健康的一个重要标准。根据《中国成人体质指数分类的推荐意见简介》[8]，一般情况下，我国成年人体质指数(BMI)在 18.5～23.9 内属于正常范围，低于 18.5 表示体重过低，高于 23.9 表示超重或肥胖，具体来讲，肥胖的标准为 BMI $\geqslant$ 28.0，超重的标准为 28$>$BMI $\geqslant$ 24。

研究高校学生 BMI 与年龄的关系，对于抑制高校学生肥胖现象的发生，具有重要的实际意义[9]。所谓 BMI 触发年龄，就是体重激增，个体达到超重或者肥胖的临界年龄。其中个体达到超重的临界年龄称为超重触发年龄，个体达到肥胖的临界年龄称为肥胖触发年龄。

现有文献表明，很多学者研究 BMI 与年龄的关系采用分位数回归的方法[10]。这种方法的缺点是无法得到总体均值函数(population mean function)和个体偏差函数(subject-specific deviation function)。

为了克服上述方法的不足，本文采用回归样条线混合效应模型(MERS 模型)的方法，研究 BMI 与年龄的关系。

MERS 模型是非参数混合效应模型的一种[11]，与分位数回归方法相比较，MERS 模型采用真实的纵向数据建模，可以得到总体均值函数和个体偏差函数。

2 MERS 模型理论

回归样条线混合效应模型(the mixed effect regression splines model)，简称 MERS 模型，具有如下的形式[12-13]：

$$y_{ij}=\eta(t_{ij})+v_i(t_{ij})+\varepsilon_{ij}$$

[收稿日期] 2023－08－28

[基金项目] 教育部人文社会科学研究规划基金项目(17YJA790098)；教育部高校思想政治工作创新发展中心 2022 年度专项研究课题(WHDHSZZX2022079)

[第一作者] 郑江松(1991－)，男，湖北英山人，湖北工业大学讲师，研究方向为产业经济学、高校思想政治教育。

$$j=1,2,\cdots,n_i;i=1,2,\cdots,n,$$
$$\varepsilon_i=[\varepsilon_{i1},\varepsilon_{i2},\cdots\varepsilon_{ini}]^{\mathrm{T}}\sim N(0,R_i) \tag{2}$$

其中 $\eta(t)$ 称为固定效应函数；$v_i(t)$ 称为随机效应函数；ε_{ij} 称为测量误差函数，y,t 分别代表个体的 BMI 和对应的年龄。

本文中 MERS 模型的基使用截断多项式基[14]。

$\eta(t)$ 的估计值使用含有 K 个结点（$\tau_1,\tau_2,\cdots,\tau_K$）的 k 阶截断多项式基 $\Phi_p(t)$，$v_i(t)$ 的估计值使用含有 K_v 个结点（$\delta_1,\delta_2,\cdots,\delta_{Kv}$）的 k_v 阶截断多项式基 $\Psi_q(t)$，即

$$\eta(t)=\boldsymbol{\Phi}_p(t)^{\mathrm{T}}\boldsymbol{\beta}=\sum_{r=0}^{k}\beta_r t^r+\sum_{l=1}^{K}\beta_{k+l}(t-\tau_l)_+^k,$$
$$v_i(t)=\boldsymbol{\Psi}_q(t)^{\mathrm{T}}b_i=\sum_{r=0}^{kv}\boldsymbol{b}_{ir}t^r+\sum_{l=1}^{Kv}b_{i(kv+l)}(t-\delta_l)_+^k, \tag{3}$$

其中 $w_+^k=\max(0,w)$

$$\boldsymbol{\beta}=[\beta_0,\beta_1,\cdots,\beta_{k+K}]^{\mathrm{T}}$$
$$\boldsymbol{b}_i=[b_{i0},b_{i1},\cdots,b_{i(kv+Kv)}]^{\mathrm{T}},i=1,2,\cdots n \tag{4}$$

记

$$\boldsymbol{x}_{ij}=\Phi_p(t_{ij}),\boldsymbol{z}_{ij}=\Psi_q(t_{ij}),$$
$$\boldsymbol{X}_i=[x_{i1},\cdots,x_{ini}]^{\mathrm{T}},\boldsymbol{y}_i=[y_{i1},\cdots,y_{ini}]^{\mathrm{T}},$$
$$\boldsymbol{Z}_i=[z_{i1},\cdots,z_{ini}]^{\mathrm{T}},\ \boldsymbol{y}=[y_1^{\mathrm{T}},\cdots y^{\mathrm{T}}],$$
$$\boldsymbol{X}=[X_1^{\mathrm{T}},\cdots,X_n^{\mathrm{T}}]^{\mathrm{T}},\boldsymbol{Z}=\mathrm{diag}(Z_1,\cdots,Z_n)$$
$$b=[b_1,\cdots,b_n]^{\mathrm{T}},\tilde{\boldsymbol{D}}=\mathrm{diag}(D,D,\cdots,D),$$
$$\boldsymbol{R}=\mathrm{diag}(R_1,\cdots,R_n),\boldsymbol{\varepsilon}=[\varepsilon_1^{\mathrm{T}},\cdots,\varepsilon_n^{\mathrm{T}}]^{\mathrm{T}} \tag{5}$$

那么式(2)就可以用矩阵表示为以下的形式：

$$\boldsymbol{y}=\boldsymbol{X\beta}+\boldsymbol{Zb}+\varepsilon$$
$$\boldsymbol{b}\sim N(0,\tilde{\boldsymbol{D}}),\varepsilon\sim N(0,R) \tag{6}$$

由文献[12]可知：以下的估计成立：

$$\hat{\boldsymbol{\beta}}=(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{X})-(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{y}),$$
$$\hat{\boldsymbol{b}}=\tilde{\boldsymbol{D}}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}), \tag{7}$$

其中

$$\boldsymbol{V}=\mathrm{diag}(V_1,\cdots V_n)$$
$$\boldsymbol{V}_i=\boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i^{\mathrm{T}}+\boldsymbol{R}_i \tag{8}$$

如果 D 和 R 未知，就通过 EM 算法[15]迭代来得到 D 和 R 的估计值。

在所有结点 $\eta(t)$ 的拟合值为：

$$\hat{\boldsymbol{\eta}}=\boldsymbol{X}\hat{\boldsymbol{\beta}}=\boldsymbol{Ay} \tag{9}$$

其中

$$\boldsymbol{A}=\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}^{-1} \tag{10}$$

在所有结点 $v_i(t)$ 的拟合向量为：

$$\hat{\boldsymbol{v}}=\boldsymbol{Z}\hat{\boldsymbol{b}}=\boldsymbol{Z}\tilde{\boldsymbol{D}}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})=$$
$$\boldsymbol{Z}\tilde{\boldsymbol{D}}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{V}^{-1}(\boldsymbol{I}_N-\boldsymbol{A})\boldsymbol{y}=\boldsymbol{A}_v\boldsymbol{y} \tag{11}$$

其中

$$\boldsymbol{A}_v=\boldsymbol{Z}\tilde{\boldsymbol{D}}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{V}^{-1}(\boldsymbol{I}_N-\boldsymbol{A}) \tag{12}$$

3　高校学生超重和肥胖触发年龄的 MERS 模型分析

3.1　数据来源及处理

本文使用的数据来源于中国健康和营养调查（China Health and Nutrition Survey，简称 CHNS）产生的数据。该调查是中国疾病预防控制中心营养与食品安全所与美国北卡罗来纳大学人口中心合作的项目。该项调查起始于 1989 年，截至目前已完成了 9 次。本文的研究对象是年龄在 9～25 岁之间的人群，从 CHNS 的有效数据中，总共得到 2171 个年龄在 9～25 岁之间个体的 7338 个观测值，由此可知，每个被抽到的个体观察的次数约为3.38 次。在所得到的样本中（样本容量为 2171），肥胖个体有 33 个，超重个体有 266 个，肥胖个体占样本的 1.5%，超重个体占样本的 12.25%。

本文的所有计算和统计分析都是使用 R 语言处理（程序清单略）。

3.2　结点选择与基函数个数的选择

运用等间距样本分位数方法（equally spaced sample quantiles as knots method）选择结点[16]，采用大多数学者常用的方法，取 $k=3$，即选取三次回归样条线，由于每个个体平均被观测了 3.38 次，为了避免奇异矩阵的出现，必须 $q\leqslant 2$，因此 $K_v=0$。

选择 p,q 时必须同时兼顾拟合优度和模型的复杂性，通常选取的标准是 AIC，其表达式为：

$$AIC=-2Loglik+2[df+df_v] \tag{13}$$

其中

$$df=tr(A),df_v=tr(A_v) \tag{14}$$

按照 p,q 的不同组合，调用 R 语言 lme4 程序包，得到图 1。

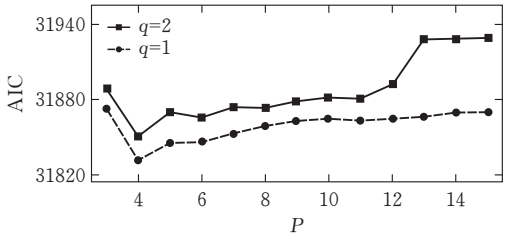


图 1　p,q 不同组合的 AIC 值

根据图 1 可知：当 $p=4,q=1$ 时 AIC 最小。

3.3　BMI 总体均值函数及其导数估计值的计算

由(3)知道 $\eta(t)$ 的 MERS 估计值为：

$$\hat{\eta}(t)=\boldsymbol{\Phi}_p(t)^{\mathrm{T}}\hat{\boldsymbol{\beta}} \tag{15}$$

由(15)知道 $\eta(t)$ 导数的 MERS 估计值：

$$\hat{\eta}(t)=\boldsymbol{\Phi}_p'(t)^{\mathrm{T}}\hat{\boldsymbol{\beta}} \tag{16}$$

调用 R 语言 lme4 程序包的 lmer 函数，得到图 2。根据图 2 可知，BMI-年龄曲线显示 9～25 岁之

间，BMI 总体均值随着年龄的增长而增加；从 BMI 导数-年龄曲线知道，在不同的阶段，BMI 随年龄增加的快慢不同，具体来讲，9～13.7 岁之间（对应于小学三年级至初中二年级阶段），BMI 总体均值增长的速率逐渐增大；13.7～21.1 岁之间（对应于初中三年级至大学本科四年级阶段），BMI 总体均值增长的速率逐渐减小；21.1～25 岁之间（对应于硕士研究生阶段），BMI 总体均值增长的速率再次逐渐增大。因此 13.7 岁时 BMI 增长的速率达到峰值（最大值），也就是说 13.7 岁（对应于初中二年级的学生）是 BMI 增速最快的年龄。

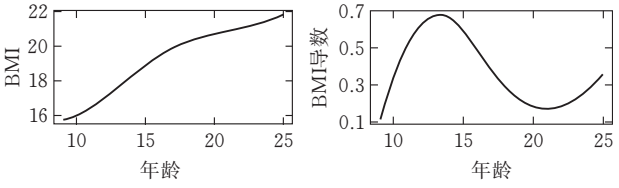


图 2 总体均值函数及其导数随年龄变化曲线图

3.4 肥胖和超重的个体函数 $s_i(t)$ 的估计值 $\hat{s}_i(t)$ 的计算

由式(2)可知，肥胖和超重个体的 BMI 个体函数 $s_i(t)=\eta(t)+v_i(t)+\varepsilon_i$，其估计值为：

$$\hat{s}_i(t)=\hat{\eta}(t)+\hat{v}_i(t) \tag{17}$$

为了得到肥胖和超重触发年龄的估计，调用 R 语言 lme4 程序包的 predict 函数，得到图 3、图 4 和图 5。

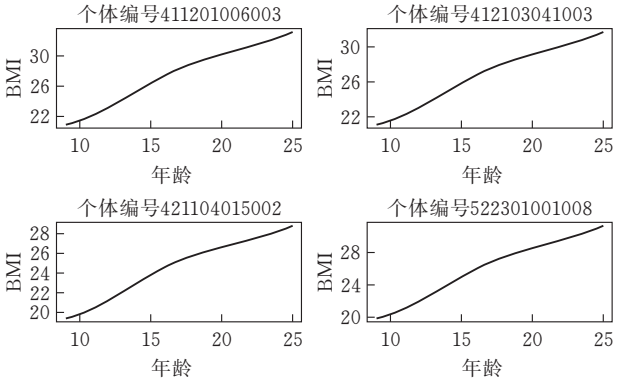


图 3 任意抽取的部分肥胖个体的 BMI-年龄曲线图

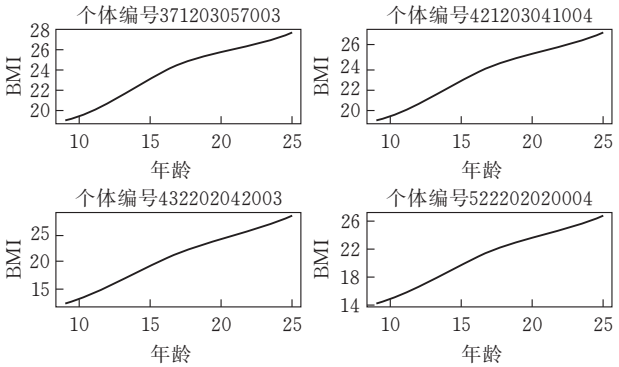
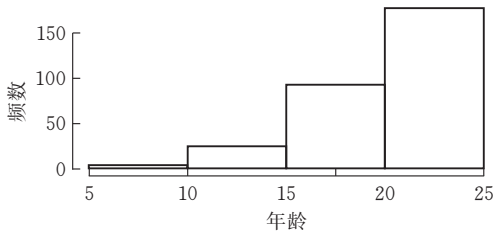
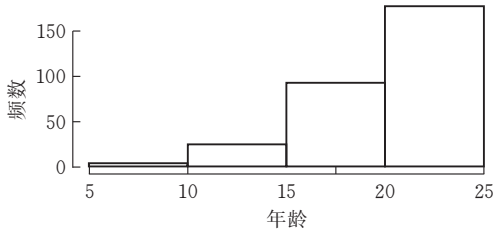


图 4 任意抽取的部分超重个体的 BMI-年龄曲线图



(a)超重样本触发年龄直方图



(b)肥胖样本触发年龄直方图

图 5 超重触发年龄、肥胖触发年龄直方图

将图 2 与图 3、图 4 进行比较不难得到，无论是肥胖个体的 BMI-年龄曲线图，还是超重个体的 BMI-年龄曲线图，都与总体的 BMI-年龄曲线图相似，这也说明选用 MERS 模型来研究 BMI 随年龄变化的规律是比较合理的。图 5 表明：无论是肥胖个体肥胖触发年龄，还是超重个体超重触发年龄都集中在 15～25 岁之间，其中超重触发年龄平均值为 20.41 岁，肥胖触发年龄平均值为 21.17 岁。

3.5 结论

本文运用回归样条线混合效应模型（MERS 模型），研究了我国 9～25 岁青少年 BMI 随年龄变化的规律，通过数据分析获得了以下结论：1）得到 BMI 总体均值函数及其导数的估计，并估计出 13.7 岁为 BMI 增速最快的年龄，即 BMI 增速的峰值发生在初中二年级；2）得到高校学生肥胖和超重的触发年龄平均值分别为 21.17 岁和 20.41 岁，即平均来看，超重的触发年龄发生在大学三年级，肥胖的触发年龄发生在大学四年级；3）发现大多数超重、肥胖个体的触发年龄在 15～25 岁之间，因此中学和大学是预防学生超重和肥胖的关键时期。结合高校学生肥胖和超重触发年龄平均值的估计，可以看出，大学本科阶段更是关键中的关键。

[参 考 文 献]

[1] Wang Y, Lobstein T. Worldwide trends in childhood overweight and obesity[J].International Journal of Pediatric Obesity, 2009, 1(03): 11-25.

[2] Karnik S,Kanekar A.Childhood obesity:A global public health crisis[J]. International Journal of Preventive Medicine, 2012, 3(01):1-7.

[3] 赵静. 1991 年－2011 年中国 2-18 岁儿童青少年超重

和肥胖患病率及流行趋势研究[D]. 石家庄:河北医科大学,2015:10-22.

[4] Prentice A M. Obesity and its potential mechanistic basis[J]. British Medical Bulletin, 2001, 60(01): 51-67.

[5] Reinehr T, Toschke A M. Onset of puberty and cardiovascular risk factors in untreated obese children and adolescents: a 1-year follow-up study[J]. Archives of Pediatrics & Adolescent Medicine, 2009, 163(08): 709-715.

[6] Nobles J, Radley D, Dimitri P, et al. Psychosocial interventions in the treatment of severe adolescent obesity: The SHINE Program[J]. Journal of Adolescent Health, 2016, 59(05): 523-543.

[7] 谭琪, 徐勇. 中国儿童青少年 1985—2010 年肥胖发展趋势及预测研究[J]. 中国学校卫生, 2013, 34(05): 570-572.

[8] 国际生命科学学会中国办事处中国肥胖问题工作组联合数据汇总分析协作组.中国成人体质指数分类的推荐意见简介[J].中华预防医学杂志,2001,35 (05): 62-63.

[9] WHO Expert Consultation. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies[J]. Lancet, 2004, 63(9403):157-163.

[10] 张艺宏,何仲涛,李宁,等. BMI 峰值年龄及超重肥胖触发年龄的研究-以四川省 2010 年国民体质监测样本为例[J]. 现代预防医学, 2013,40(04) :108-117.

[11] Huang J Z, Wu C O, Zhou L. Varying-coefficient models and basis function approximations for the analysis of repeated measurements[J]. Biometrika, 2002(89): 111-128.

[12] Wu H, Zhang J T. Nonparametric regression methods for longitudinal data analysis[M]. Wiley-Interscience. 2005:130-136.

[13] 王友乾, 付利亚, 徐建文. 纵向数据分析[M]. 北京:高等教育出版社, 2015.

[14] Wold S. Spline functions in data analysis[J]. Technometrics, 1974(16): 1-11.

[15] Laird N M, Ware J H. Random effects models for longitudinal data[J]. Biometrics, 1982(38): 963-974.

[16] Jerome H, Friedman, Bernard W, et al. Flexible parsimonious smoothing and additive modeling[J]. Technometrics, 1989, 31(01): 3-21.

Statistical Analysis of Trigger Ages of Overweight and Obesity Among College Students Based on MERS Model

ZHENG Jiangsong[1], WU Shibao[2], LUO Youxi[2]

(1 *School of Electrical and Electronic Engineering , Hubei Univ. of Tech., Wuhan* 430068,*China*;
2 *School of Science, Hubei Univ. of Tech., Wuhan* 430068,*China*)

Abstract: BMI (Body Mass Index) is an important index to measure overweight and obesity. The evolution of BMI with age is an important basis for obesity intervention in adolescents. The purpose of this study is to investigate the relationship between BMI and age, and to estimate the trigger age of overweight and obesity. In order to overcome the shortcomings of the existing "Quantile regression method", the authors use AIC index to select suitable node number and smoothing parameters according to the mixed effect regression splines model (MERS model). The following conclusions have been obtained: (1) the estimation of population mean function and the derivative of population mean function has been obtained, and the age of 13.7 years is estimated to be the fastest growing age of BMI; (2) the individual function of obese and overweight individuals is estimated, and the average trigger ages of obese and overweight individuals are 21.17 and 20.41 years respectively; (3) the most of overweight and obese individuals are found to be triggered between the ages of 15 and 25, so high school and college are critical periods for preventing overweight and obesity in students, especially college students are in the most critical periods for preventing overweight and obesity.

Keywords: BMI; MERS model; overweight; obesity; trigger age

[责任编校: 赵　瞳]