

基于有序聚类分析法探讨成年人体质指数的年龄和性别分布特征

蔡亚宁 裴晓婷 孙盼盼 徐亦平 刘莉 平智广

450001 郑州大学公共卫生学院卫生统计学教研室

通信作者:平智广, Email: ping_zhg@163.com

DOI: 10.3760/cma.j.issn.0254-6450.2018.06.024

【摘要】 目的 探讨我国成年人BMI的年龄和性别分布特征。方法 采用中国健康与营养调查(CHNS)2009年的调查数据,选取 ≥ 20 岁且年龄、身高、体重资料完整的成年人作为研究对象,分别按5、10岁分组后运用SAS 9.2软件对各组平均BMI再按性别进行有序聚类分析。结果 有序聚类分析将我国成年人BMI值按年龄特征划分为3类,5岁分组的性别合并组及女性组聚类结果第一类为20~岁,第二类为40~65岁,第三类为 > 65 岁;5岁分组男性组第一类为20~岁,第二类为40~60岁,第三类为 > 60 岁。10岁分组中,所有分组结果一致,均为第一类20~岁,第二类为40~60岁,第三类为 > 60 岁。提示5岁分组比10岁分组聚类分析结果更精细,BMI的性别分布特征之间存在一定差异。结论 我国成年人BMI按年龄特征可划分为3类,与“老中青”3个年龄段的年龄划分基本一致,且青年人群随年龄增长,BMI呈上升趋势,中年人群BMI约稳定在 24 kg/m^2 ,老年人群BMI呈下降趋势。

【关键词】 体质指数; 年龄; 性别; 有序样品聚类分析

基金项目: 国家自然科学基金(81001280, 81202277); 河南省高等学校重点科研项目(16A330003)

Using the sequenced sample cluster analysis to study the body mass index distribution characteristics of adults in different age groups and genders Cai Yaning, Pei Xiaoting, Sun Panpan, Xu Yiping, Liu Li, Ping Zhiguang

Department of Health Statistics, School of Public Health, Zhengzhou University, Zhengzhou 450001, China

Corresponding author: Ping Zhiguang, Email: ping_zhg@163.com

【Abstract】 Objective To explore the characteristics of distribution on Chinese adult body mass index (BMI) in different age groups and genders and to provide reference related to obesity and related chronic diseases. **Methods** Data from the China Health and Nutrition Survey in 2009 were used. Sequential sample cluster method was used to analyze the characteristics of BMI distribution in different age groups and genders by SAS. **Results** Our results showed that the adult BMI in China should be divided into 3 groups according to their age, as 20 to 40 years old, 40 to 65 years old, and > 65 years old, in females or in total when grouped by difference of 5 years. For groupings in male, the three groups should be as 20 to 40, 40 to 60 years old and > 60 years old. There were differences on distribution between the male and female groups. When grouped by difference of 10 years, all of the clusters for male, female and total groups as 20-40, 40-60 and > 60 years old, became similar for the three classes, respectively, with no differences of distribution between gender, suggesting that the 5-years grouping was more accurate than the 10-years one, and BMI showing gender differences. **Conclusions** BMI of the Chinese adults should be divided into 3 categories according to the characteristics of their age. Our results showed that BMI was increasing with age in youths and adolescents, remained unchanged in the middle-aged but decreasing in the elderly.

【Key words】 Body mass index; Age; Gender; Sequenced sample cluster

Fund programs: National Natural Science Foundation of China (81001280, 81202277); Henan Province Higher Education Key Research Funding Project (16A330003)

BMI是衡量人体胖瘦程度及营养状况的常用指标,并与肺活量、握力等人体体质指标有着显著联

系^[1],BMI超标还是心血管疾病等肥胖相关疾病的危险因素^[2-3]。BMI变化是一个连续性过程,考虑

到生长发育连续性的特点,研究时不能把年龄随意地分成各年龄组。有序样品聚类法考虑到连续性,有助于正确评估年龄对 BMI 的影响。为此本研究采用中国健康与营养调查 (China Health and Nutrition Survey, CHNS) 2009 年的数据,采用有序样品聚类分析的方法,根据我国成年人 BMI 和性别对年龄进行分组,探讨 BMI 在不同年龄组和性别间的分布特征及其变化趋势。

对象与方法

1. 研究对象:源自 CHNS 2009 年数据^[4]。该调查采用多阶段分层整群抽样方法在 9 个省份抽取 2 个城市和 4 个县,共抽取 18 个市和 36 个县,在城市分别抽取 2 个城市居委会和 2 个郊区村,在县分别抽取 1 个县政府所在的居委会和 3 个村,最后在抽取的所有居委会和村中随机抽取 20 户居民,为覆盖各年龄段,将户内每个家庭成员作为调查对象。调查内容涵盖人口学特征、健康状况等^[5-6]。按统一标准测量调查对象身高、体重^[7],并计算 BMI(kg/m²)。采用实足年龄的算法,即调查当日的日期减去被调查者的出生日期,出生日期记为阴历者换算成阳历后再计算,精确到小数点后两位。本研究选取 ≥20 岁且年龄、身高、体重资料完整的成年人作为研究对象。

2. 有序样品的聚类—最优分割法^[8]:该法基于方差分析,寻找 1 个分割,使各分段内样本间的差异最小而各分段之间样本的差异最大。

(1) 定义类直径 $D(i, j)$:设变量 x_1, x_2, \dots, x_n 的某一类是 $\{X_i + X_{i+1} + \dots + X_j\}$, $j > i$, 定义其均值向量为

$$\bar{X}_{ij} = (X_i + X_{i+1} + \dots + X_j) / (j - i + 1) \quad (1)$$

将该类内部各样本间的总差异 (其指标是离差平方和) 定义为该类的直径,用 $D(i, j)$ 表示

$$D(i, j) = \sum_{l=i}^j (X_l - \bar{X}_{ij})^2 \quad (2)$$

(2) 定义误差函数 $L[p(n, k)]$:设将 n 个有序变量分为类,某一分割为 $p(n, k): \{i_1, i_1 + 1, \dots, i_2 - 1\}, \{i_2, i_2 + 1, \dots, i_3 - 1\}, \dots, \{i_k, i_k + 1, \dots, n\}$, 定义这种分类的误差函数即目标函数为类内总离差平方和

$$L[p(n, k)] = \sum_{j=1}^k D(i_j, i_{j+1} - 1) \quad (3)$$

(3) 确定聚类数:当 n, k 固定时,误差函数 $L[p(n, k)]$ 越小表示各类的类内离差平方和越小,分类越合理。最优分割就是使 $L[p(n, k)]$ 达到最小值时的一种分类法,聚类数 k 通过做 $L[p(n, k)]$ 与聚类数 k 的变化趋势图求得,曲线拐点处的 k 值即为最优聚类数。

本研究中有序样本 BMI 进行年龄聚类分析,就是保持年龄的顺序性不变,找出分界点将 BMI 样本科学分类。将 n 个有序变量 BMI 分为 k 类,各类的误差函数为 $L[p(n, k)]$ 。当 n, k 固定时, L_p 越小,分类越合理。聚类数 k 通过做误差函数 L_p 与聚类数 k 的变化趋势图求得,曲线拐点处的 k 值即为最优聚类数。

3. 统计学分析:所有研究对象分别按 5、10 岁为一个年龄段分组,运用 SAS 9.2 软件统计男、女性以及合计人口的各年龄段 BMI 均值;绘制 10、5 岁分组及性别合并组 BMI 均值折线统计图,对年龄段分组的 BMI 均值做有序聚类分析。

结 果

1. 一般特征:选取男性 3 912 人,女性 4 471 人,合计 8 383 人作为研究对象。以 5 岁为一个年龄段,将研究对象分为 20~、25~、30~、...、85~、90~、≥95 岁组,分为标记为第 1, 2, 3, ..., 16 组,男、女性及其性别合并组的 BMI 均值统计见表 1。以 10 岁为一个年龄段,将研究对象分为 20~、30~、...、80~、≥90 岁组,标记为第 1, 2, ..., 7, 8 组,男、女性及其性别合并组的 BMI 均值统计见表 2。由于 5 岁分组中 ≥80 岁人数较少予以合并,因此分为 13 个组;10 岁分组中 ≥70 岁人数较少予以合并,分为 6 个组。

表 1 成年人按 5 岁年龄段分组的 BMI 均值(kg/m²)

年龄组 (岁)	组号	男性		女性		合并	
		人数	BMI 均值	人数	BMI 均值	人数	BMI 均值
20~	1	180	21.70	185	21.12	365	21.41
25~	2	159	22.84	199	21.58	358	22.14
30~	3	232	23.46	264	22.25	496	22.81
35~	4	382	23.61	420	23.02	802	23.30
40~	5	409	23.79	526	23.61	935	23.69
45~	6	458	24.24	528	24.06	986	24.14
50~	7	477	23.49	556	24.04	1 033	23.79
55~	8	506	23.51	559	24.06	1 065	23.80
60~	9	372	23.14	412	24.00	784	23.59
65~	10	299	23.31	292	23.90	591	23.60
70~	11	197	23.09	277	23.53	474	23.34
75~	12	156	22.30	151	23.03	307	22.66
80~	13	67	21.71	74	22.04	141	21.88
85~	14	16	21.60	22	22.78	38	22.28
90~	15	2	20.39	5	20.87	7	20.74
95~	16	0	0.00	1	17.37	1	17.37

2. 各组 BMI 均值的变化趋势:5、10 岁分组性别合并组两组曲线变化大致相同。随着年龄增长,从 20 岁起 BMI 均值不断增加,约 40 岁时 BMI 均值曲

线保持平稳,65岁后BMI均值总体呈现下降趋势(图1)。

表2 成年人按10岁年龄段分组的BMI均值(kg/m²)

年龄组(岁)	组号	男性		女性		合并	
		人数	BMI均值	人数	BMI均值	人数	BMI均值
20~	1	339	22.24	384	21.36	723	21.77
30~	2	614	23.55	684	22.72	1 298	23.12
40~	3	867	24.03	1 054	23.84	1 921	23.92
50~	4	983	23.50	1 115	24.05	2 098	23.79
60~	5	671	23.21	704	23.96	1 375	23.60
70~	6	353	22.74	428	23.35	781	23.07
80~	7	83	21.69	96	22.21	179	21.97
90~	8	2	20.39	6	20.29	8	20.31

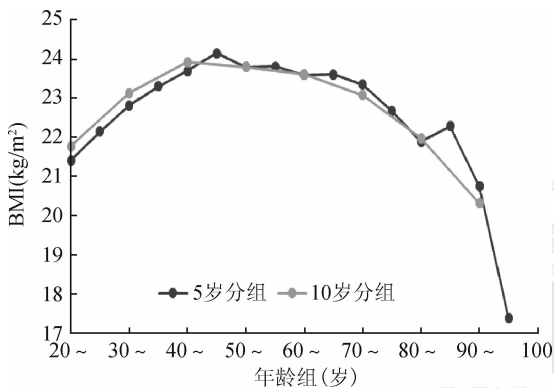


图1 成年人按5、10岁分组的BMI均值(kg/m²)年龄变化趋势

3. 有序聚类分析:以5岁分组男性为例,最小损失函数出现在聚类数为3和4时,说明年龄组被分为3或4类效果较好(图2)。当聚类数 $k=3$ 时,查图3男性聚类结果显示:第一类为年龄组别1~4组,第二类为年龄组别5~8组,第三类为年龄组别9~13组,即第一类:20~岁,第二类:40~60岁,第三类:>60岁。当聚类数 $k=4$ 时,第一类:20~岁,第二类:35~岁,第三类:50~70岁,第四类:>70岁。结合实际情况,分为3类更具有意义。因此,5岁分组男性聚类结果应为:第一类:20~岁,第二类:40~60岁,第三类:>60岁。其他分组结果分析过程与此相同;10岁分组的不同分类数所对应的损失函数见图4,聚类结果见图5。所有分组聚类结果汇总为表3。

讨论

BMI值过大是罹患心脑血管疾病、糖尿病等肥胖相关疾病的危险因素^[9],过小则会增加感染性疾病死亡的危险性^[10]。研究表明随年龄增长,BMI值呈连续性变化^[11]。因此,确定BMI指标有序聚类点对了解人体健康发展特征,制定科学合理有针对性的相关慢性病防治措施具有重要意义。

本研究采用有序聚类分析的结果显示,按5岁

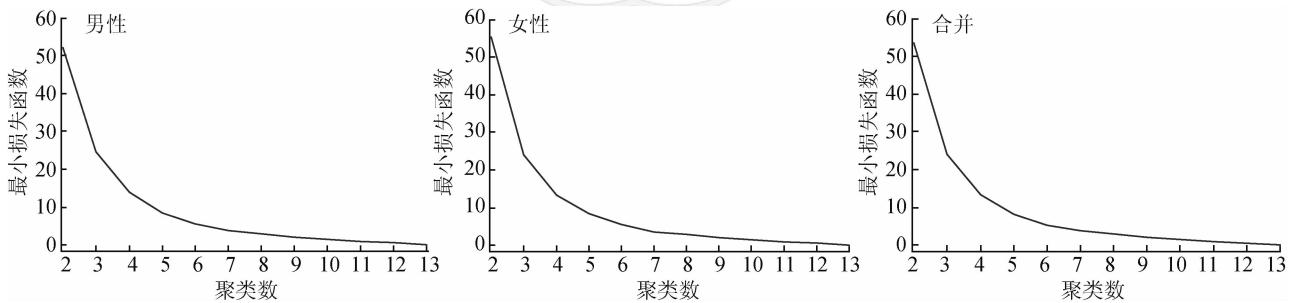


图2 5岁分组不同分类数最小损失函数

男性		女性		合并	
k	组号	k	组号	k	组号
1		1		1	
2	1 8	2	1 7	2	1 7
3	1 5 9	3	1 5 10	3	1 5 10
4	1 4 7 11	4	1 4 7 11	4	1 4 8 11
5	1 3 5 6 9 12	5	1 4 6 9 12	5	1 3 6 9 12
6	1 3 5 7 9 12	6	1 3 5 7 10 12	6	1 3 5 7 9 12
7	1 2 4 6 8 10 12	7	1 3 5 7 9 11 13	7	1 2 4 6 8 10 12
8	1 2 3 5 7 9 11 12	8	1 3 4 5 7 9 11 13	8	1 2 4 6 8 10 12 13
9	1 2 3 5 7 9 11 12 13	9	1 3 4 5 6 8 10 12 13	9	1 2 3 4 6 8 10 12 13
10	1 2 3 5 6 7 9 11 12 13	10	1 2 3 4 5 6 8 10 12 13	10	1 2 3 4 6 7 9 11 12 13
11	1 2 3 5 6 7 9 10 11 12 13	11	1 2 3 4 5 6 8 10 11 12 13	11	1 2 3 4 5 6 7 9 11 12 13
12	1 2 3 4 5 6 7 9 10 11 12 13	12	1 2 3 4 5 6 7 9 10 11 12 13	12	1 2 3 4 5 6 7 8 9 11 12 13
13	1 2 3 4 5 6 7 8 9 10 11 12 13	13	1 2 3 4 5 6 7 8 9 10 11 12 13	13	1 2 3 4 5 6 7 8 9 10 11 12 13

注:k表示聚类数,组号表示聚类的相应起始年龄组,例如男性中, $k=2$ 时聚类为两类,第一类年龄组从1开始,第二类从8开始;即第一类为年龄组1~7,第二类为年龄组8~13;分3类时,第一类为年龄组1~4,第二类为年龄组5~8,第三类为年龄组9~13,依次类推

图3 5岁分组的聚类结果

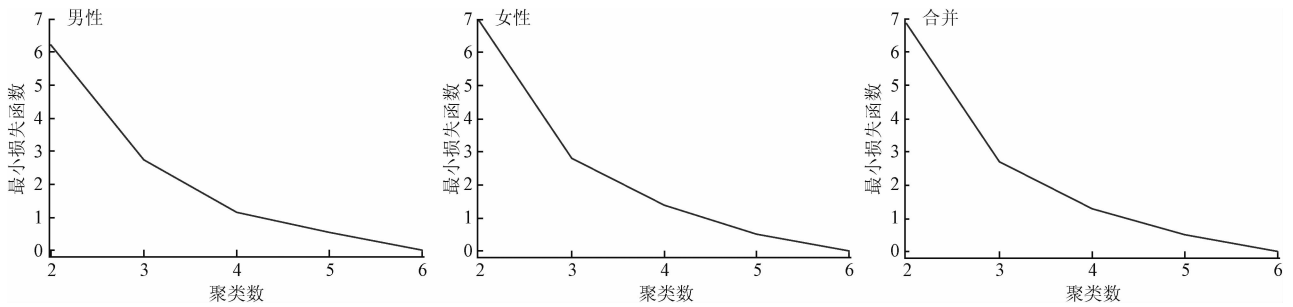


图4 10岁分组不同分类数最小损失函数

男性		女性		合并	
k	组号	k	组号	k	组号
1		1		1	
2	1 4	2	1 3	2	1 4
3	1 3 5	3	1 3 5	3	1 3 5
4	1 2 4 6	4	1 2 3 5	4	1 2 3 5
5	1 2 3 4 6	5	1 2 3 4 6	5	1 2 3 5 6
6	1 2 3 4 5 6	6	1 2 3 4 5 6	6	1 2 3 4 5 6

注:k表示聚类数,组号表示聚类的相应起始年龄组

图5 10岁分组的聚类结果

表3 5和10岁分组的聚类结果汇总

序号	分组	聚类分组号	年龄组(岁)聚类结果
1	5岁分组男性	1~4 5~8 9~13	20~ 40~60 >60
2	5岁分组女性	1~4 5~9 10~13	20~ 40~65 >65
3	5岁分组合并	1~4 5~9 10~13	20~ 40~65 >65
4	10岁分组男性	1~2 3~4 5~6	20~ 40~60 >60
5	10岁分组女性	1~2 3~4 5~6	20~ 40~60 >60
6	10岁分组合并	1~2 3~4 5~6	20~ 40~60 >60

分组时男性20~岁为第一类,第二类为40~60岁,第三类为>60岁;女性组与性别合并组聚类结果相同,第一类为20~岁,第二类为40~65岁,第三类为>65岁。10岁分组中,所有聚类结果一致,第一类为20~岁,第二类为40~60岁,第三类为>60岁。按年龄特征有序聚类将我国成年人BMI划分为3类,与“老中青”3个年龄段的年龄划分基本一致,且这3类BMI变化特征不同。第一类(20~岁组,青年人群)BMI均值随着年龄增长呈增加趋势,罹患肥胖相关疾病的危险性已明显增加^[12],应进行肥胖高危因素的早期干预,控制体重,改善随年龄增长BMI值趋高现状。第二类(40~65岁,中年人群)BMI均值达到高峰且稳定在24 kg/m²左右,达到我国人体超重标准,即“肥胖前期”^[13],是多种肥胖相关疾病的高发年龄段,应作为控制体重、预防肥胖相关慢性疾病的关键时期。第三类(>60或>65岁,老年人群)的BMI呈现下降趋势且变化幅度在3类之中最为明显。这是因为老年人群随着年龄增长,机体组织发生改变、器官逐渐老化,随之而来的是适应能力减弱,抵抗免疫力降低,导致身体素质随年龄增长呈明

显下降^[14];也可能与该数据为横断面调查,中年时期BMI值过大的人群其生存时间可能较短^[15],可能存在老年人群的结果偏倚。总之,在慢性疾病的防制中应考虑不同年龄段人群的生理差异,结合BMI值的变化特征制定不同年龄、性别的卫生政策。

国际上对老年人界定标准有65岁或60岁。前者多为发达国家采纳,后者则被大多数发展中国家所接受。但从BMI的角度进行年龄聚类时,结果存在差异,即对于第三类聚类点是60岁还是65岁的确定,5岁分组和10岁分组性别合并组不同、5岁分组中男、女性结果也存在分歧。两种结果呈现的差异可能的原因:①分组间隔不同:10岁年龄分组比较粗,掩盖了以5岁分组为聚类点的变化;②在我国人口老龄化加速发展时期,在老年人口中女性所占比例大于男性^[16]。由此本研究认为5岁分组结果可能更为准确,BMI的聚类特点在男、女性别间也存在差异。当前我国经济发展及老龄化速度加速,人群预期寿命^[17]、年龄组成^[18]也在发生变化,本文男、女老龄组的划分界值差异可能与经济和人口构成存在联系,且数据为CHNS 2009年结果,随着我国经济的发展,老年期的年龄界限可能会达到65岁。

本研究采用有序聚类方法充分考虑了人群BMI的数据结构,有效反映了BMI内在分布规律,准确将人群按照生理状况的差异进行归类,从而能够更全面反映不同年龄段群体间BMI变化,其中BMI的有序聚类恰好与我国“老中青”年龄分组相一致,更具有参考价值。

本研究中仅采用BMI指标,而无与肥胖相关疾病亦有密切联系的血压、血脂等指标。因此有必要进一步扩大其他涉及的指标,以明确不同年龄人群身体机能与肥胖及相关慢性疾病的关系。

利益冲突 无

参 考 文 献

- [1] 程小游,钟新生. BMI分布规律及不同BMI人群体质水平研究[J]. 福建体育科技, 2014, 33(4): 27-30. DOI: 10.3969/j.issn.1004-8790.2014.04.009.
Cheng XY, Zhong XS. Study on the BMI distribution and the physical fitness about different BMI [J]. Fujian Sports Sci Technol, 2014, 33(4): 27-30. DOI: 10.3969/j.issn.1004-8790.2014.04.009.
- [2] Huotari PRT, Nupponen H, Laakso L, et al. Secular trends in muscular fitness among Finnish adolescents [J]. Scand J Public Health, 2010, 38(7): 739-747. DOI: 10.1177/1403494810384425.
- [3] 王悠清. 成人体质指数与死亡风险的前瞻性研究[D]. 杭州: 浙江大学, 2015.
Wang YQ. Association between body mass index and mortality in adults: A prospective study in China [D]. Hangzhou: Zhejiang University, 2015.
- [4] Fang C, Liang Y. Social disparities in body mass index (BMI) trajectories among Chinese adults in 1991-2011 [J]. Int J Equity Health, 2017, 16(1): 146. DOI: 10.1186/s12939-017-0636-5.
- [5] 刘爱东,张兵,王惠君,等. 1997-2009年中国九省区成人体质指数分布及变化趋势研究[J]. 中国健康教育, 2011, 27(3): 167-170. DOI: 10.16168/j.cnki.issn.1002-9982.2011.03.007.
Liu AD, Zhang B, Wang HJ, et al. Distribution of body mass index and its changing trends among Chinese adults in nine provinces from 1997 to 2009 [J]. Chin J Health Educat, 2011, 27(3): 167-170. DOI: 10.16168/j.cnki.issn.1002-9982.2011.03.007.
- [6] China Health and Nutrition Survey. Acknowledgement [EB/OL]. (2011) [2017]. <http://www.cpc.unc.edu/projects/china/data/acknowledgement.html>.
- [7] Song P, Li X, Gasevic D, et al. BMI, waist circumference reference values for Chinese school-aged children and adolescents [J]. Int J Environ Res Public Health, 2016, 13(6): 589. DOI: 10.3390/ijerph13060589.
- [8] 朱建平. 应用多元统计分析[M]. 2版. 北京: 科学出版社, 2012.
Zhu JP. Applied multivariate statistical analysis [M]. 2nd ed. Beijing: Science Press, 2012.
- [9] 中国肥胖问题工作组数据汇总分析协作组. 我国成人体重指数和腰围对相关疾病危险因素异常的预测价值: 适宜体重指数和腰围切点的研究[J]. 中华流行病学杂志, 2002, 23(1): 5-10. DOI: 10.3760/j.issn:0254-6450.2002.01.003.
Cooperative Meta-analysis Group of China Obesity Task Force. Predictive values of body mass index and waist circumference to risk factors of related diseases in Chinese adult population [J]. Chin J Epidemiol, 2002, 23(1): 5-10. DOI: 10.3760/j.issn:0254-6450.2002.01.003.
- [10] 王俊,高玉堂,王学励,等. 上海市中老年男性体重指数与死亡的前瞻性研究[J]. 中华流行病学杂志, 2005, 26(6): 394-399. DOI: 10.3760/j.issn:0254-6450.2005.06.003.
Wang J, Gao YT, Wang XL, et al. A prospective cohort study on body mass index and mortality among middle-aged and elderly men in urban Shanghai [J]. Chin J Epidemiol, 2005, 26(6): 394-399. DOI: 10.3760/j.issn:0254-6450.2005.06.003.
- [11] 刘亮. 山西省成年人超重和肥胖变化趋势的研究[D]. 太原: 中北大学, 2015.
Liu L. Adults overweight and obesity trends research in Shanxi Province Abstract [D]. Taiyuan: North University of China, 2015.
- [12] 梁柳娟,黄朝褪,张婉芬,等. 社区成年人肥胖流行病学调查[J]. 现代预防医学, 2003, 30(6): 774-775, 778. DOI: 10.3969/j.issn.1003-8507.2003.06.010.
Liang LJ, Huang ZT, Zhang WF, et al. Epidemiological survey on obesity among adults in community [J]. Mod Prev Med, 2003, 30(6): 774-775, 778. DOI: 10.3969/j.issn.1003-8507.2003.06.010.
- [13] 武阳丰,周北凡,陶寿淇,等. 我国中年人群超重率和肥胖率的现状及发展趋势[J]. 中华流行病学杂志, 2002, 23(1): 11-15. DOI: 10.3760/j.issn:0254-6450.2002.01.004.
Wu YF, Zhou BF, Tao SQ, et al. Prevalence of overweight and obesity in Chinese middle-aged populations: Current status and trend of development [J]. Chin J Epidemiol, 2002, 23(1): 11-15. DOI: 10.3760/j.issn:0254-6450.2002.01.004.
- [14] 刘德鹤. 江西省60~69岁老年人体质现状以及对比分析[D]. 南昌: 江西师范大学, 2013.
Liu DH. The current situation and comparative analysis of 60-69 old people's physical fitness of Jiangxi [D]. Nanchang: Jiangxi Normal University, 2013.
- [15] Odegaard AO, Pereira MA, Koh WP, et al. BMI, all-cause and cause-specific mortality in Chinese Singaporean men and women: the Singapore Chinese Health Study [J]. PLoS One, 2010, 5(11): e14000. DOI: 10.1371/journal.pone.0014000.
- [16] Jia YZ, Tan L. A study on the feminization of old population in China [J]. Chin Populat Today, 2012, 58(5): 38.
- [17] 段白鸽. 我国全年龄段人口平均预期寿命的动态演变[J]. 人口与经济, 2015, (1): 49-63. DOI: 10.3969/j.issn.1000-4149.2015.01.005.
Duan BG. Dynamic evolution of all ages life expectancies in China [J]. Populat Economics, 2015, (1): 49-63. DOI: 10.3969/j.issn.1000-4149.2015.01.005.
- [18] 张建,华琦. 中国老龄化的特征发展趋势与对策[J]. 中国心血管杂志, 2012, 15(1): 79-80. DOI: 10.3969/j.issn.1007-5410.2010.01.029.
Zhang J, Hua Q. The characteristics of China's aging trend and countermeasures [J]. Chin J Cardiovascu Med, 2012, 15(1): 79-80. DOI: 10.3969/j.issn.1007-5410.2010.01.029.

(收稿日期: 2017-10-11)

(本文编辑: 张林东)