

# adaptive LASSO logistic 回归模型应用于老年人 养老意愿影响因素研究的探讨\*

韩耀风<sup>1,2</sup> 覃文峰<sup>1,2</sup> 陈 炜<sup>1,2</sup> 李博涵<sup>1,2</sup> 滕伯刚<sup>1,2</sup> 方 亚<sup>1,2,Δ</sup>

**【提 要】** 目的 探讨 adaptive LASSO logistic 回归模型在老年人养老意愿影响因素研究中的应用。方法 基于厦门市 60 岁及以上老年人口的多阶段整群抽样调查数据,建立老年人养老意愿影响因素的 adaptive LASSO logistic 回归模型,通过交叉验证法选择模型中的调和参数  $\lambda$ ;通过与全变量和逐步 logistic 回归结果的比较,探讨 adaptive LASSO logistic 回归模型的优势。结果 共纳入 1244 名老年人,其养老意愿为家庭养老、社区居家养老和机构养老的比例分别为 70.0%、21.1% 和 8.9%。交叉验证法选择的  $\lambda$  为 0.018;此时 adaptive LASSO logistic 回归模型纳入的自变量为居住地、年龄、婚姻状况、文化程度、子女数、每月退休金收入、公费医疗和住院情况;BIC 和 AIC 分别为 1931、1888,均低于全变量 logistic 回归(2077、1923)和逐步 logistic 回归(2025、1912)。结论 adaptive LASSO logistic 回归模型可用于老年人养老意愿影响因素研究。老年人的养老意愿受多个因素影响。

**【关键词】** adaptive LASSO logistic 回归模型 养老模式 影响因素

## Influencing Factors on Elder's Preference for Supporting: Application of an adaptive LASSO logistic Model

Han Yaofeng, Qin Wenfeng, Chen Wei, et al (School of Public Health, Xiamen University, 361102, Xiamen)

**【Abstract】 Objective** This study was aimed to analyze influencing factors on elder's preference for supporting in case to provide evidence for diversified elderly supporting. **Methods** This article applied an adaptive LASSO logistic model to the multi-stage cluster sampling data of the population aged 60 or older in Xiamen to analyze influencing factors on elder's preference for supporting. Cross validation method was used to choose  $\lambda$  for adaptive LASSO logistic model. In addition, we evaluated the model fitting of adaptive LASSO logistic model by comparing the BIC and AIC with full logistic model and stepwise logistic model. **Results** The cross validation method resulted in  $\lambda = 0.018$  for adaptive LASSO logistic model, in which variables retained were residence, age, marital status, education level, number of children, the monthly pension income, medical insurance and hospitalization. BIC of adaptive LASSO logistic model, full logistic model and stepwise logistic model were 1931, 2077 and 2025 respectively. And AIC of the three models were 1888, 1923 and 1912 respectively. **Conclusion** Compared to full logistic model and stepwise logistic model, adaptive LASSO logistic model was the best fitting model for influencing factors on elder's preference for supporting. Adaptive LASSO logistic model could be used to analyze influencing factors on elder's preference for supporting. There were multiple factors which influenced elder's preference for supporting.

**【Key words】** Adaptive LASSO logistic model; Elderly supporting; Influencing factors

2010 年第六次全国人口普查结果显示,中国 60 岁及以上老龄人口已达到 1.78 亿,所占比例从 2000 年的 10.2% 增至 13.3%<sup>[1]</sup>。随着老龄化速度的不断加快,在经济发展与社会结构转型的过程中,传统家庭养老和单纯的机构养老远远不能满足当今社会的养老需求。在此背景下社区居家养老应运而生,形成了多种养老方式并存的多元化养老模式。老年人养老意愿的影响因素复杂,如何选择自变量是研究的关键点。本研究拟基于对厦门市老年人养老意愿的调查数据,探讨 adaptive LASSO logistic 回归模型在老年人养老意愿影响因素分析研究中的应用,分析老年人养老意愿的影响因素,为完善多元化养老模式提供依据。

### 基本原理

#### 1. LASSO logistic 回归模型

对于多变量 logistic 回归问题,一般常使用变量子集选择的方法来选择变量。Zou 等人<sup>[2]</sup>提出变量子集选择有两点限制:首先,当变量的个数较多时,子集选择的运算量庞大,甚至可能无法运算;其次,因为子集选择方法存在内在离散性<sup>[3]</sup>,子集选择的结果是不稳定的<sup>[4]</sup>。

作为变量子集选择的替代方案,逐步回归克服了变量较多时的运算量问题,但是得到的结果仍具有内在离散性和不稳定性<sup>[3-4]</sup>。同时逐步回归得到的结果一般为局部最优解而不是全局最优解<sup>[2]</sup>,且忽略了变量选择过程中的随机误差和不确定性<sup>[5-6]</sup>。

LASSO 方法<sup>[6-8]</sup>用模型的绝对系数函数作为惩罚项来压缩模型的系数,使绝对值较小的系数为 0,达到同时进行变量选择和参数估计的目的,而传统方法

\* 基金项目:厦门大学大学生创新创业训练计划项目(2015Y0827)

1. 厦门大学公共卫生学院(361102)

2. 卫生技术评估福建省高校重点实验室

Δ通信作者:fangya@xmu.edu.cn

变量选择和参数估计是分开进行的。LASSO 方法很好地克服了逐步回归变量选择方法的局限性,同时又保留了子集选择和岭回归的优良性质。采用 LASSO 方法选择自变量而建立的 logistic 回归模型即为 LASSO logistic 回归模型。

假设有独立同分布的观测值  $(X^i, y_i)$   $i = 1, 2, \dots, n$ , 其中  $X^i = (x_{i1}, \dots, x_{ip})$  和  $y_i$  是模型的自变量和因变量。logistic 回归模型的条件概率如式(1)所示。

$$\ln \left\{ \frac{p(y_i = 1 | X^i)}{1 - p(y_i = 1 | X^i)} \right\} = \eta_\beta(X^i) \quad (1)$$

其中

$$\eta_\beta(X^i) = \beta_0 + \sum_{j=1}^p x_{ij} \beta_j \quad (2)$$

LASSO logistic 回归模型中的系数估计值  $\hat{\beta}$  则可以写成式(3)的形式:

$$\hat{\beta} = \operatorname{argmin} \sum_{i=1}^n \{ y_i \eta_\beta(X^i) - \ln\{1 + \exp[\eta_\beta(X^i)]\} + \lambda \sum_{j=1}^p |\beta_j| \} \quad (3)$$

其中  $\lambda$  是一个非负的正则化参数,它决定了 LASSO logistic 回归模型系数的压缩程度,称为调和参数。随着  $\lambda$  的增大,各个变量系数估计值逐渐被压缩,当  $\lambda$  足够大时,一些变量系数被压缩为 0。研究表明, LASSO 方法具有优良的理论性质,适合于稀疏的多变量回归问题<sup>[9-11]</sup>。选择  $\lambda$  的常用方法有自助法、交叉验证法、广义交叉验证法等。在交叉验证法中,每个子样本验证一次,交叉重复  $n$  次,得到  $n$  次的模型拟合情况,从中选择最优  $\lambda$ 。

### 2. adaptive LASSO logistic 回归模型

LASSO logistic 模型虽然可以较好地地进行多变量的选择,但得到的结果是有偏估计,不具有模型的相合性和参数估计渐进正态性,即不具有 Oracle Property (哲人性质)<sup>[2,4]</sup>。Zou 等人于 2006 年提出了一个改进的 LASSO 方法,被称为 adaptive LASSO 方法<sup>[2]</sup>。adaptive LASSO logistic 回归模型既具有 LASSO logistic 回归模型的优点,同时估计结果具有渐进无偏性和 Oracle Property。

adaptive LASSO 方法的改进之处在于,进行参数估计时对于不同的系数分配了不同的权重。令  $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p)$  为全模型的最大似然估计值,以式(4)为权重,则 adaptive LASSO logistic 回归模型中的系数估计值  $\hat{\beta}$  为式(5)。

$$\hat{\omega}_j = \frac{1}{|\hat{\beta}_j|^\gamma} \quad (\gamma > 0) \quad (4)$$

$$\hat{\beta} = \operatorname{argmin} \sum_{i=1}^n \{ y_i \eta_\beta(X^i) - \ln\{1 + \exp[\eta_\beta(X^i)]\} + \lambda \sum_{j=1}^p \hat{\omega}_j |\beta_j| \} \quad (5)$$

## 资料与方法

### 1. 资料来源

采用横断面调查的方法,调查厦门市 60 岁及以上常住老年人,调查内容包括个人信息(性别、年龄、子女数、婚姻情况、文化程度、居住地)、生活基本情况(每月退休金、城镇职工基本医疗保险、城镇居民基本医疗保险、公费医疗)、健康状况(健康自评、是否曾住院)和养老意愿(家庭养老、社区居家养老、机构养老)等信息。

### 2. 统计分析

采用 Epidata 3.1 软件进行双人双录入资料,使用 R 3.2.0 进行统计分析。本研究使用卡方检验分析分类变量与养老意愿的关系,通过建立 adaptive LASSO logistic 回归模型分析老年人养老意愿的影响因素,其中  $\lambda$  的选择使用交叉验证法。将 adaptive LASSO logistic 模型与全变量 logistic 回归、逐步 logistic 回归进行比较,采用 AIC 和 BIC 准则评估 adaptive LASSO logistic 回归模型的拟合效果。检验水准取  $\alpha = 0.05$ 。

## 结果

### 1. 基本情况

本次调查共发放问卷 1328 份,回收有效问卷 1244 份,有效应答率为 93.67%。1244 名老年人的年龄 60~100 岁,中位年龄 71 岁;其养老方式选择意愿的分布分别为家庭养老(70.0%)、社区居家养老(21.1%)、机构养老(8.9%)。各分析变量及养老方式意愿情况见表 1。

### 2. 老年人养老方式选择的单因素分析

对各分类变量与养老方式选择进行单因素卡方检验,结果见表 1。其中居住地、文化程度、每月退休金、城镇职工基本医疗保险、城镇居民基本医疗保险、公费医疗、是否曾住院等因素有统计学意义( $P < 0.05$ )。

### 3. 老年人养老方式选择的多因素分析

#### (1) 变量选择与模型估计

在单因素分析的基础上,以居住地、年龄、子女数、文化程度、每月退休金、城镇职工基本医疗保险、城镇居民基本医疗保险、公费医疗、是否曾住院等因素作为自变量,建立 logistic 回归模型,分析它们对养老方式选择的影响,各变量的赋值情况见表 2。

以家庭养老为参照,利用 R 软件包 glmnet<sup>[12]</sup> 进行 adaptive LASSO logistic 回归模型分析。通过交叉验证得到模型误差与  $\lambda$  的关系,折叠次数为 10 次,结果如图 1 示。由图 1 可见,模型误差最小时对应的  $\lambda$  为 0.018,此时入选的变量有居住地、年龄、婚姻状况、文化程度、子女数、每月退休金、参加公费医疗和住院情况,即这些因素为老年人养老方式选择的影响因素。

表 1 各分析变量的养老方式意愿情况

	养老方式			$\chi^2$	P
	家庭养老(%)	社区居家养老(%)	机构养老(%)		
性别					
男(n=1052)	69.8	21.3	8.9	0.233	0.8900
女(n=192)	71.4	19.8	8.9		
婚姻状况					
在婚(n=833)	68.5	22.9	8.5	5.405	0.0670
非在婚(n=411)	73.0	17.3	9.7		
文化程度					
文盲(n=539)	73.8	19.9	6.3	52.121	<0.0001
小学(n=308)	76.0	18.2	5.8		
初中(n=192)	66.1	24.0	9.9		
高中(n=113)	58.4	25.7	15.9		
大学及以上(n=92)	50.0	26.1	23.9		
居住地					
城镇(n=573)	68.1	18.8	13.1	29.164	<0.0001
农村(n=671)	72.0	23.4	4.6		
每月退休金(元)					
<1000(n=679)	73.2	21.5	5.3	30.134	<0.0001
1000~(n=105)	75.2	14.3	10.5		
≥2000(n=460)	67.0	18.4	14.5		
城镇职工基本医疗保险					
是(n=828)	71.1	21.9	7.0	11.401	0.0030
否(n=416)	67.8	19.5	12.7		
城镇居民基本医疗保险					
是(n=654)	72.2	22.0	5.8	16.485	0.0003
否(n=590)	67.6	20.0	12.4		
公费医疗					
是(n=61)	47.5	34.4	18.0	16.034	0.0003
否(n=1183)	71.2	20.4	8.5		
曾住院					
是(n=207)	60.4	23.7	15.9	17.804	0.0001
否(n=1037)	71.9	20.5	7.5		
健康自评					
好(n=559)	72.1	20.4	7.5	3.730	0.4440
一般(n=573)	68.6	21.1	10.3		
差(n=112)	67.0	24.1	8.9		

表 2 变量赋值表

变量	符号	赋值
养老方式	Y	家庭养老: Y=0, 社区居家养老: Y=1, 机构养老: Y=2
居住地	D1	农村: D1=0 城镇: D1=1
年龄(岁)	X1	X1 ∈ [60, 100]
性别	D2	男: D2=0 女: D2=1
子女数	X2	X2 ∈ N <sup>0</sup>
文化程度	D3	文盲: D3=0, 小学: D3=1, 初中: D3=2, 高中: D3=3, 大学及以上: D3=4
月退休金(元)	D4	<1000: D4=0, 1000~: D4=1, ≥2000: D4=2
城镇职工基本医疗保险	D5	否: D5=0 是: D5=1
城镇居民基本医疗保险	D6	否: D6=0 是: D6=1
公费医疗	D7	否: D7=0 是: D7=1
曾住院	D8	否: D8=0 是: D8=1
健康自评	D9	好: D9=0, 一般: D9=1, 差: D9=2
婚姻状况	D10	非在婚: D10=0 在婚: D10=1

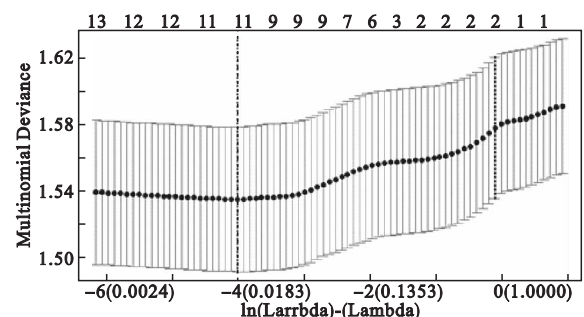


图 1 Lambda 与模型误差

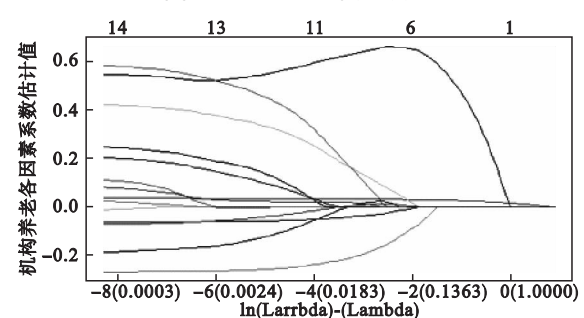


图 2 Lambda 与变量的解路径

\* : 上方的数字为模型变量选择数; 横坐标( ) 内的数字为 λ

图 2 为随着 λ 值的变化模型变量的筛选情况。可

见 随着  $\lambda$  增大 模型压缩程度增大 模型中包含的自变量个数减少 模型选择主要变量的功能增强。

(2) 模型的参数估计

为了对比 LASSO 变量选择的结果 ,本研究还建立了全变量 logistic 回归模型和逐步 logistic 回归模

型。各模型的参数估计及 *AIC* 和 *BIC* 结果见表 3 ,其中 adaptive LASSO logistic 回归模型的 *AIC* 和 *BIC* 均为最小 ,而全变量 logistic 回归的 *AIC* 和 *BIC* 均为最大。

表 3 不同回归模型参数估计结果

	全变量 logistic 回归		逐步 logistic 回归		adaptive LASSO logistic 回归模型	
	社区居家养老	机构养老	社区居家养老	机构养老	社区居家养老	机构养老
( Intercept)	-2. 684 **	-6. 957 ***	-2. 357 **	-6. 565	-2. 192	-6. 389
居住地	-0. 750 **	0. 447	-0. 622 ***	0. 493	-0. 635	0. 573
年龄( 岁)	0. 022*	0. 065 ***	0. 023*	0. 064	0. 019	0. 060
婚姻状况	0. 357*	0. 067	0. 366*	0. 072	0. 311	0. 133
曾住院	-0. 281	-0. 542*	-0. 291	-0. 534	-0. 282	-0. 496
文化程度						
小学	-0. 120	0. 010	-0. 99	0. 008	0. 000	-0. 000
初中	0. 343	0. 492	0. 379	0. 521	0. 180	0. 141
高中	0. 442	0. 866*	0. 527	0. 913*	0. 418	0. 588
大学及以上	0. 443	1. 112**	0. 532	1. 148**	0. 403	0. 683
子女数	-0. 105	-0. 144	-0. 113	-0. 144	-0. 099	-0. 126
月退休金( 元) 1000 ~	-0. 274	0. 257	-	-	-0. 050	0. 019
≥2000	0. 094	0. 194	-	-	0. 184	0. 121
职工保险	0. 540	0. 272	-	-	-	-
居民保险	0. 354	0. 374	-	-	-	-
公费医疗	1. 181*	0. 328	0. 750*	0. 073	0. 596	0. 214
<i>BIC</i>	2077		2025		1931	
<i>AIC</i>	1923		1912		1888	

\* :  $P < 0. 1$  , \*\* :  $P < 0. 05$  , \*\*\* :  $P < 0. 01$  .

讨 论

对于多变量选择问题 ,子集选择具有运算量大和变量选择结果不稳定的缺点。逐步回归克服了较多变量时子集选择的运算量问题 ,但是变量子集选择具有内在的不连续性 ,从而导致变量子集选择的结果不稳定 ,逐步回归方法得到的变量选择结果并不一定是最佳子集<sup>[6]</sup>。逐步回归方法还忽略了变量选择过程中的随机误差和不确定性。adaptive LASSO logistic 模型使用绝对系数函数作为惩罚项来压缩模型的系数 ,同时在参数估计时对于不同的系数分配了不同的权重。它既具有 LASSO logistic 回归模型的优点 ,同时参数的估计结果具有渐进无偏性和 Oracle Property。adaptive LASSO logistic 回归模型很好地克服了传统变量选择方法在选择模型方面的不足 ,同时又保留了子集选择和岭回归的优良性质 ,适合于多个影响因素研究中的变量选择问题。本研究以厦门市 60 岁及以上常住老年人的养老意愿数据为基础 ,拟合全变量 logistic 回归、逐步 logistic 回归和 adaptive LASSO logistic 回归模型。结果发现 ,adaptive LASSO logistic 回归模型的 *AIC* 值和 *BIC* 值均小于全变量 logistic 回归、逐步 logistic 回归 ,说明 adaptive LASSO logistic 回归模型对数据的拟合效果最好。

老年人养老方式选择的影响因素较为复杂 ,往往

受多种因素的影响。本研究的单因素分析、adaptive LASSO logistic 回归模型、全变量 logistic 回归模型和逐步 logistic 回归模型均显示 ,居住地、年龄、文化程度、公费医疗和住院情况与养老方式的选择有关联。在本次调查中 ,城市老年人选择家庭养老的比例高于农村 ,而农村老年人选择机构养老的比例低于城镇;赵海林等人对淮安市的研究和谭小林等对重庆市老年人的调查也发现这样的差异性<sup>[13-14]</sup>。本研究发现随着年龄的增长 ,老年人选择家庭养老的可能性降低。刘小春等人对江西省的研究发现 ,年龄越长者越倾向于选择机构养老模式 ,年龄越小者越倾向于选择家庭养老模式或社区居家养老模式<sup>[15]</sup>。但陈玉兰等的研究发现 ,老年人的年龄越小 ,越倾向选择机构养老和社区居家养老<sup>[19]</sup>。国内多数学者均发现 ,文化程度是老年人养老意愿的影响因素 ,文化程度越高 ,选择机构养老和社区居家养老的比例越高 ,而文化程度越低 ,选择家庭养老的比例越高<sup>[15-17]</sup>。本研究还发现享有公费医疗的老年人选择机构养老和社区居家养老的比例高于未享有公费医疗的 ,而后者选择家庭养老的比例高于前者;曾住院的老年人选择机构养老和社区居家养老的比例高于未曾住院的 ,而后者选择家庭养老的比例高于前者。目前很少有学者关注公费医疗、住院情况与养老意愿的关系。本研究的 adaptive LASSO logistic 回归模型、全变量 logistic 回归和逐步 logistic 回归

模型结果显示婚姻状况与老年人的养老方式选择有关,在婚的老年人选择社区居家养老的比例较高,而非在婚的老年人选择家庭养老和机构养老的比例较高。国内多项研究发现,婚姻状况与老年人的养老意愿有关<sup>[15,17]</sup>。

子女数、每月退休金收入仅在 adaptive LASSO logistic 回归模型中显示出与老年人养老意愿的相关性。随着老年人的子女数增多,与社区居家养老、机构养老相比,老人更有可能选择家庭养老,说明子女对于老年人的抚养对老年人养老方式的选择有一定的影响。顾永红等人对全国 9 个省份 27 个地级市进行的研究和刘小春等人的研究均显示子女数与老年人养老意愿之间的这种关系<sup>[15,18]</sup>。选择机构养老意味着老年人需要定期给养老机构支付一定数额的资金,需要一定的经济基础。本研究发现,每月养老金收入高于 2000 元的老年人,更有可能选择机构养老或社区居家养老。覃丹等人以湖南省某市开展的研究<sup>[19]</sup>和赵海林等人的研究<sup>[13]</sup>也得出了类似的结论,研究发现经济有保障和社会地位较高老年人群,退休金一般在 2000 元以上,对生活质量 and 精神文化生活要求较高,在生活方面比较独立,能够自费购买老年服务或者选择入住养老机构,在对养老模式的选择上更具有主动性<sup>[19]</sup>。

综上所述,adaptive LASSO logistic 回归模型可用于老年人养老方式选择影响因素分析,得到的模型具有更好的解释性。老年人养老意愿受多个因素的影响,应加强农村、高龄、在婚、低文化程度、子女数较多、低收入、未享有公费医疗和未曾住院老年人的养老保障的覆盖及社会养老的宣传教育,开发有针对性的养老方式。

#### 参 考 文 献

- [1] 中华人民共和国国家统计局. 2010 年第六次全国人口普查主要数据公报(第 1 号). 中国计划生育学杂志 2011 54(8):511-512.
- [2] Zou H. The adaptive lasso and its oracle properties. Journal of the American statistical association 2006 101(476):1418-1429.
- [3] Breiman L. Better subset regression using the nonnegative garrote.

Technometrics 1995 37(4):373-384.

- [4] Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American statistical Association 2001 96(456):1348-1360.
- [5] Shen X, Ye J. Adaptive model selection. Journal of the American Statistical Association 2002 97(457):210-221.
- [6] 张秀秀, 王慧, 田双双, 等. 高维数据回归分析中基于 LASSO 的自变量选择. 中国卫生统计 2013 30(6):922-926.
- [7] Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society 1996 58(1):267-288.
- [8] 陈江鹏, 彭斌, 文雯, 等. 微阵列数据中的先验信息对基于 LASSO 变量选择方法影响的模拟研究. 中国卫生统计 2015 32(3):407-409.
- [9] Meinshausen N, Bühlmann P. High-dimensional graphs and variable selection with the lasso. The annals of statistics 2006 34(3):1436-1462.
- [10] Donoho DL, Elad M. Optimally sparse representation in general( non-orthogonal) dictionaries via  $\ell_1$  minimization. Proceedings of the National Academy of Sciences 2003 100(5):2197-2202.
- [11] Donoho DL, Huo X. Uncertainty principles and ideal atomic decomposition. Information Theory, IEEE Transactions on 2001 47(7):2845-2862.
- [12] Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. Journal of statistical software 2010 33(1):1.
- [13] 赵海林. 淮安市老年人养老模式选择意愿实证分析. 南京人口管理干部学院学报 2012 28(2):9-12.
- [14] 方匡南, 章贵军, 张惠颖. 基于 Lasso-logistic 模型的个人信用风险预警方法. 数量经济技术经济研究 2014(2):125-136.
- [15] 刘小春, 李婵. 中部欠发达地区城镇居民养老模式选择行为实证分析——基于江西省的调查数据. 社会保障研究 2014(2):24-31.
- [16] 陶涛, 丛聪. 老年人养老方式选择的影响因素分析——以北京市西城区为例. 人口与经济 2014(3):15-22.
- [17] 扈映, 杨康, 舒泰. 农村居民养老居住意愿选择的实证研究. 调研世界 2014(8):24-29.
- [18] 顾永红. 农村老年人养老模式选择意愿的影响因素分析. 华中师范大学学报(人文社会科学版) 2014 53(3):9-15.
- [19] 覃丹. 城市老年人养老需求及养老模式选择研究. 华中师范大学, 2015.

(责任编辑:郭海强)

(上接第 17 页)

- [9] Stone M. Cross-validated choice and assessment of statistical predictions. Journal of the royal statistical society. Series B( Methodological) , 1974 36(2):111-147.
- [10] Szymańska, Saccenti E, Smilde AK, et al. Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. Metabolomics 2012 8(1):3-16.

- [11] Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. Appears in the International Joint Conference on Artificial Intelligence( IJCAI) ,1995.
- [12] Hall P, Racine J, Li Q. Cross-validation and the estimation of conditional probability densities. J Am Stat Assoc 2004 9(2):115-126.

(责任编辑:郭海强)