

· 应用研究 ·

基于 LASSO 和分位数回归的人均卫生费用影响因素分析

谢 聪^{1#} 黄淑琼^{1#} 温昊于² 宇传华^{2△} 李明炎^{1△}

【摘要】目的 筛选并分析我国人均卫生费用影响因素。**方法** 采用 LASSO 方法和分位数回归模型对 2018 年全国 31 个省、直辖市、自治区的人均卫生总费用的影响因素进行分析。**结果** LASSO 方法从 45 个协变量中筛选出 8 个对我国各省人均卫生费用最重要的影响变量。在最终筛选出的 8 个变量中,孕产妇死亡率和每千人口卫生技术人员数显著促进各省人均卫生费用,且系数估计值随人均卫生费用的上升呈现先增加后减小的特征。居民消费支出的系数估计值为正,系数估计值除了 70%分位点没有统计学意义,在其他四个分位点均有统计学意义。死亡率、公立医院中政府办医院数、三级医院数、门诊健康检查人数、居民人均可支配收入的系数估计值在不同分位数有正有负。**结论** 不同分位的人均卫生费用水平,各影响因素效应各异。对于人均卫生费用水平较低的省份,应关注居民消费支出的影响;对于人均卫生费用水平中等的省份,应关注孕产妇死亡率和每千人口卫生技术人员的影响。在增加卫生总费用投入的同时,适当控制卫生总费用的增长速度,实现卫生事业与国民经济均衡发展。

【关键词】 LASSO 分位数回归 人均卫生费用 影响因素

【中图分类号】 R197.1

【文献标识码】 A

DOI 10.11783/j.issn.1002-3674.2024.05.014

准确了解我国人均卫生费用的影响因素及影响机制能够为制定精准控费政策提供理论依据^[1-3]。从现有的研究文献中不难发现,影响各省的人均卫生费用的因素众多,大部分研究选取自变量时偏向主观,这容易造成某些影响显著的自变量被遗漏,或者引入影响不显著的自变量,从而导致实证模型的估计和预测准确度大大降低。而 LASSO 估计法较好地克服了传统变量选择方法的不足,权衡了回归模型的精确性和可解释性两个重要指标,在实际应用中得到大量有效应用^[4]。另外,现有的相关研究更多使用均值回归探究不同因素对卫生费用的影响,但均值回归常常忽略某些重要信息。相比传统线性回归而言,分位数回归在不同分位数水平上进行分析,可以得到更加全面丰富的结果,适用于任何分布尤其是非正态分布的数据资料^[5]。

鉴于上述,本文将运用 LASSO 估计法,客观准确地筛选出真正影响我国人均卫生费用的因素。再采用分位数回归模型分析筛选后的影响因素对卫生费用的影响机制,更好地帮助研究者和决策者深入了解卫生费用的影响因素,有助于其结合实际情况制定精准控费政策。

资料方法

1. 数据来源与变量设计

本研究使用的数据来自于《中国统计年鉴》和《中国卫生统计年鉴》,主要包括 2018 年全国 31 个省、直辖市、自治区的人均卫生总费用等指标。根据已有文

献,本研究基于经济因素、卫生资源、卫生投入、医疗服务、人口因素和人群健康状况六个维度,选取了 45 个可能影响地区人均卫生费用水平的变量,见表 1。

2. 研究方法

1996 年,Robert Tibshirani 提出 LASSO 回归模型,即 Least Absolute Shrinkage and Selection Operator^[6],它的基本思想是在回归系数的绝对值之和小于一个常数的约束条件下,使残差平方和最小化。这一方法兼顾了子集选择的可解释性和岭回归的稳定性。

LASSO 方法是在普通线性回归模型中增加惩罚项,对于普通线性模型 LASSO 估计为:

$$\text{Lasso} = \arg \min_{\beta \in R^d} |Y - X\beta|^2 \text{ s.t. } \sum_{j=1}^d |\beta_j| \leq t, t \geq 0$$

等价于:

$$\text{Lasso} = \arg \min_{\beta \in R^d} (|Y - X\beta|^2 + \lambda \sum_{j=1}^d |\beta_j|)$$

其中 t 与 λ 一一对应,为调节参数。

令 $t_0 = \sum_{j=1}^d |\hat{\beta}_j(\text{OLS})|$, 当 $t < t_0$ 时,一部分系数就会被压缩为 0,从而降低 X 的维度,达到减小模型复杂度的目的。通过对调节参数 λ 的控制,可实现变量筛选。

由于本文所选取的 45 个自变量中,它们的量纲具有不一致性,因此本研究对这些自变量数据进行标准化处理后再进行 LASSO 回归。

分位数回归是 Koenker 和 Bassett 提出的一种回归估计方法,是一种基于因变量的条件分布来拟合自变量线性函数的回归模型^[7]。分位数回归优点主要是对异常值更加稳健,并且可以同时研究自变量对模型位置参数和尺度参数的影响,从而使得对数据有更

1. 湖北省疾病预防控制中心预防医学信息研究所(430079)

2. 武汉大学公共卫生学院

共同第一作者

△通信作者:宇传华,E-mail:yuchua@163.com;李明炎,E-mail:limingyan@163.com

加全面和准确的理解,且适合于异方差数据。分位数回归模型如下:

$$Q(\beta_q) = \sum_{i: y_i \geq x'_i \beta} q |y_i - x'_i \beta| + \sum_{i: y_i < x'_i \beta} (1 - q) |y_i - x'_i \beta|$$

表 1 可能影响地区人均卫生费用水平的变量

影响因素	指标名称
卫生资源	每千人口卫生技术人员数(人)、基层医疗机构数(个)、基层医疗机构床位数(张)、基层医疗机构人员数(人)、公立医院中政府办医院数(个)、公立医院数(个)、医疗卫生机构数(个)、三级医院数(个)、二级医院数(个)、一级医院数(个)、中医类医疗机构床位数(个)、中医类医疗卫生机构数(个)
卫生投入	一般公共预算收入(亿元)、地方一般公共预算支出(亿元)、一般公共预算支出中的医疗卫生与计划生育支出(亿元)、职工基本医保基金收入(亿元)
医疗服务	居民年住院率(%)、平均住院日(日)、家庭卫生服务人次数(人次)、医师日均担负诊疗人次(人次)、住院病人手术人次数(人次)、居民平均就诊次数(次)、每百门诊入院人数(人)、病床使用率(%)、医院分科门诊人次数(万人次)、门诊健康检查人数(万人)、入院人数(人)、出院人次数(人次)、病床工作日(日)
人口因素	65岁及以上占比(%)、0~14岁占比(%)、出生率(%)、少年儿童抚养比(%)、老年人口抚养比(%)、城镇人口比重(%)、文盲率(%)
人群健康状况	2010年预期寿命(岁)、死亡率(‰)、甲乙类法定报告传染病发病率(1/10万)、甲乙类法定报告传染病死亡率(1/10万)、孕产妇死亡率(1/10万)
经济因素	人均地区生产总值(元)、居民消费支出(千元)、居民消费支出的医疗保健支出(元)、居民人均可支配收入(千元)

研究结果

1. 利用 LASSO 回归模型筛选变量结果

本研究的 LASSO 方法由 R 语言的 glmnet 软件包完成,通过交叉验证得到模型误差与 λ 的关系,迭代次数为 100 次。通过计算,均方误差最小时对应的 λ 为 0.041(见图 1)。图 2 中的垂线为交叉验证法所确定的最优的 λ,此时与该垂线相交的曲线所代表的变量即为 LASSO 模型筛选出来的重要变量,从而筛选出 8 个变量,包括死亡率、孕产妇死亡率、每千人口卫生技术人员数、居民消费支出、公立医院中政府办医院数、三级医院数、门诊健康检查人数、居民人均可支配收入(见图 2),即这些因素为我国各省人均卫生费用的重要影响因素。

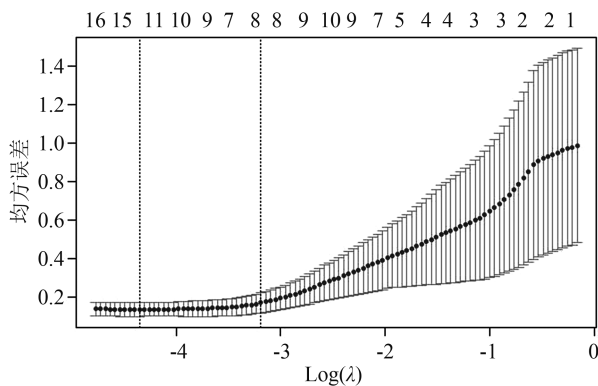
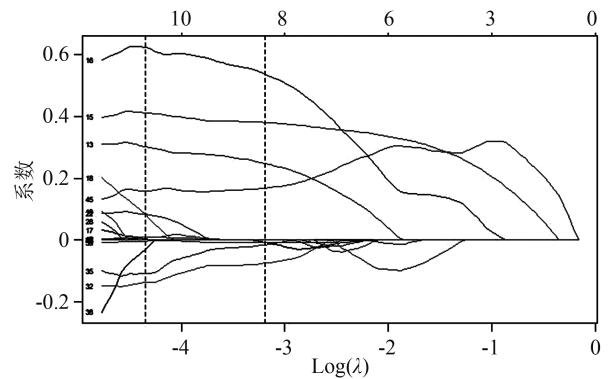


图 1 Log(λ) 与均方误差

2. 分位数回归模型的估计结果

表 2 列出了经 LASSO 回归得出的 8 个影响因素在 10%、30%、50%、70% 和 90% 五个分位点的系数估计结果,为进一步观察参数变化情况,将 8 个变量的五个分位点系数估计值绘制在图 3 中。



注:图中数字为 Lasso 回归模型选出的变量个数

图 2 LASSO 变量筛选图

从表 2 可以看出,孕产妇死亡率和每千人口卫生技术人员数对各省人均卫生费用的影响系数估计值为正,在五个分位点都有统计学意义,且系数估计值随分位数的增加呈现先增加后减小的特征。居民消费支出的系数估计值为正,系数估计值除了在 70%分位点没有统计学意义,在其他四个分位点均有统计学意义。死亡率、公立医院中政府办医院数、三级医院数、门诊健康检查人数、居民人均可支配收入的系数估计值在不同分位数有正有负,但死亡率在 10%分位数的系数估计值显著为负,三级医院数在 30%和 50%分位数的系数估计值显著为负,门诊健康检查人数在 90%分位数显著为负,居民人均可支配收入在 10%分位数显著为负,90%分位数显著为正。

讨论

以往文献都证明 LASSO 方法与逐步回归、最优子集选择、岭回归、主成分回归、偏最小二乘回归等传统变量选择方法相比,LASSO 方法筛选的变量模型更易于解释且计算相对简单方便^[8]。本文使用 LASSO

方法将许多可能影响人均卫生费用水平的因素纳入到研究中,通过这一方法筛选出主要影响因素,再纳入到分位数回归模型之中进行实证分析,一定程度上解决了传统回归模型中变量选择过于主观

的问题,并使得到的模型具有更好的可解释性,且更好的识别均值回归不能发现的影响因素,度量对因变量特定分位数的边际效果,使得各系数估计值更加稳定和精确。

表 2 人均卫生费用的分位数回归模型系数估计值

变量	分位数				
	10%	30%	50%	70%	90%
公立医院中政府办医院数	-0.41	0.21	0.33	0.98	0.36
孕产妇死亡率(1/10万)	45.77***	64.90***	65.48***	54.19***	47.88***
每千人口卫生技术人员数	588.07***	637.23***	676.47***	601.74***	540.27***
居民消费支出(千元)	328.14***	187.26***	132.52**	121.19	120.04***
三级医院数	1.97	-8.82**	-7.38**	-3.44	0.51
门诊健康检查人数(万人)	-0.30	0.36	0.24	-0.33	-0.56***
死亡率(%)	-296.63***	56.10	120.08	-92.44	-66.98
居民人均可支配收入(千元)	-130.09***	6.09	46.89	53.16	49.49**

注: ** 和 *** 分别代表在 5% 和 1% 的水平上显著。

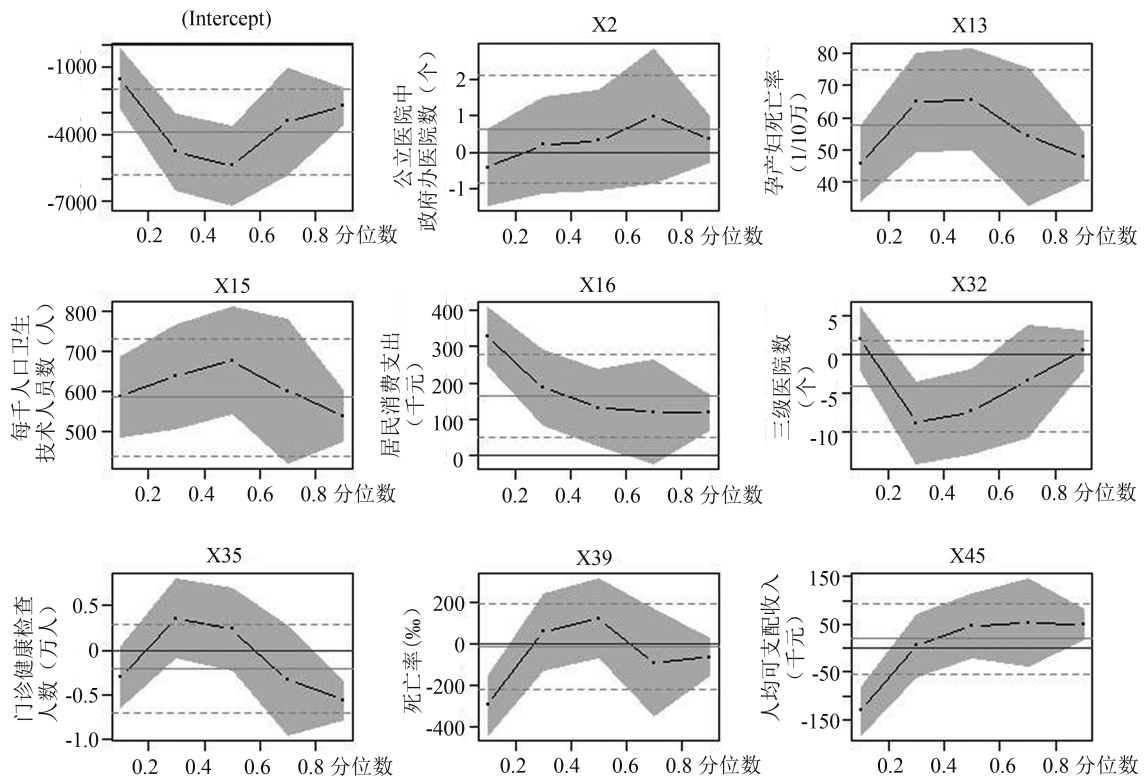


图 3 人均卫生费用在不同分位数的回归系数估计值

孕产妇死亡率的大小反映了某地妇女和儿童健康水平的高低,是人群健康状态和健康产出的指标之一。从我们的研究可以发现孕产妇死亡率对人均卫生费用是显著的正效应,与黄玉捷等的发现孕产妇死亡率与卫生总费用呈现正相关的结果一致^[9]。由于降低孕产妇死亡率对推动妇幼健康事业高质量发展,保障妇女儿童健康,推进健康中国建设具有重大意义,因此国家政府部门一直持续加大与孕产妇健康相关的卫生费用投入。根据分位数回归的结果来看,孕产妇的死亡率对人均卫生费用的影响效应在低分位点和高分位点

小于中分位点,降低孕产妇死亡率对于降低人均卫生费用水平中等地区的人均卫生费用有较强的影响。

每千人口卫生技术人员数代表着一个地区卫生人力资源的配置,对人均卫生费用呈现显著的正效应,这是因为更充足的卫生技术人员带来了更好的医疗卫生服务和更多的医疗卫生支出。分位数回归的结果显示系数先增加后减小,说明每千人口卫生技术人员数在人均卫生费用水平处于中等时产生的影响更大,因此对于人均卫生费用水平中等的地区,保证卫生技术人员的合理配置对于维持本地区卫生费用水平有重要意

义。

居民消费支出与人均卫生费用有正相关性,居民的医疗保健支出是居民消费支出的重要组成部分,一般来说居民的消费支出越高,说明居民的经济实力更强,从而能够负担更多的医疗保健支出。但随着人均卫生费用水平逐渐增加,居民消费支出的影响逐渐减小,这是由于随着居民消费支出的增加,医疗保健支出的占比并没有逐步升高。因此对于人均卫生费用水平较低的地区,增加居民的消费支出,则能大幅提高人均卫生费用水平,而对于人均卫生费用水平较高的地区,增加居民的消费支出对卫生费用的影响有限。

结 论

本研究采用 LASSO 方法和分位数回归对我国人均卫生费用影响因素进行筛选并进行实证分析,得到的模型具有更好的精确性和解释性。针对不同分位的人均卫生费用水平,各影响因素效应各异。不同分位的人均卫生费用水平,各影响因素效应各异。对于人均卫生费用水平较低的省份,应关注居民消费支出的影响;对于人均卫生费用水平中等的省份,应关注孕产妇死亡率和每千人口卫生技术人员的影响。在增加卫生总费用投入的同时,适当控制卫生总费用的增长速

度,实现卫生事业与国民经济均衡发展。

参 考 文 献

- [1] 文捷, 杜福贻, 卢祖洵, 等. 我国卫生总费用影响因素及实证研究[J]. 中国全科医学, 2016, 19(7): 824-827.
- [2] 颜琰. 我国人均卫生费用的主成分分析[J]. 中国卫生经济, 2017, 36(12): 43-45.
- [3] 张炳文, 李小菊, 毛璐, 等. 基于组合预测模型的新疆生产建设兵团个人卫生支出分析[J]. 中国卫生事业管理, 2021, 38(7): 511-514.
- [4] 张兴祥, 钟威, 洪永淼. 国民幸福感的指标体系构建与影响因素分析: 基于 LASSO 的筛选方法[J]. 统计研究, 2018, 35(11): 3-13.
- [5] 谢聪, 宇传华, 张爽等. 基于省际面板分位数回归的中国城乡居民医疗保健支出影响因素分析[J]. 中国卫生统计, 2018, 35(1): 26-28+32.
- [6] Tibshirani R. Regression shrinkage and selection via the lasso[J]. Journal of the Royal Statistical Society. Series B (Methodological), 1996, 58: 267-288.
- [7] Koenker R. Quantile regression for data longitudinal data[J]. Journal of Multivariate Analysis, 2004, 91(1): 74-89.
- [8] 陶春海, 王梦颖. 基于 Lasso 回归模型的我国个人卫生支出占比影响因素分析[J]. 统计与决策, 2017, (21), 100-103.
- [9] 黄玉捷. 不同健康产出水平下卫生总费用与健康产出对应关系研究[J]. 卫生软科学, 2019, 33(3): 16-21.
- (责任编辑: 邓 妍)
- (上接第 700 页)
- [5] Feng ZL, Liu C, Guan XP, et al. China's rapidly aging population creates policy challenges in shaping a viable long-term care system [J]. Health Affairs, 2012, 31(12): 2764-2773.
- [6] Zhu H. Unmet needs in long-term care and their associated factors among the oldest old in China[J]. BMC Geriatrics, 2015, 15: 46.
- [7] Feng J, Wang Z, Yu YY. Does long-term care insurance reduce hospital utilization and medical expenditures? Evidence from China[J]. Social Science and Medicine, 2020, 258: 113081.
- [8] Chen H, Ning J. The impacts of long-term care insurance on health care utilization and expenditure: evidence from China[J]. Health Policy and Planning, 2022, 37(6): 717-727.
- [9] Tang Y, Chen TR, Zhao Y, et al. The Impact of the Long-Term Care Insurance on the Medical Expenses and Health Status in China [J]. Frontiers in Public Health, 2022, 10: 847822.
- [10] Zhao Y, Strauss J, Chen X, et al. China Health and Retirement Longitudinal Study Wave 4 User's Guide[EB/OL]. (2020-09-01) [2022-10-01]. [https://charls.pku.edu.cn/en/data/User 2018.pdf](https://charls.pku.edu.cn/en/data/User%202018.pdf).
- [11] Zhao Y, Hu Y, Smith JP, Strauss J, Yang G. Cohort profile: the China health and retirement longitudinal study (CHARLS) [J]. International Journal of Epidemiology 2014, 43: 61-68.
- [12] Lei XY, Bai C, Hong JP, et al. Long-term care insurance and the well-being of older adults and their families: Evidence from China [J]. Social Science and Medicine, 2022, 296: 114745.
- [13] Zhu YM, Österle A. Rural-urban disparities in unmet long-term care needs in China: the role of the hukou status[J]. Social Science and Medicine, 2017, 191: 30-37.
- [14] Garcia-Gomez P, Hernandez-Quevedo C, Jimenez-Rubio D, et al. Inequity in long-term care use and unmet need: two sides of the same coin[J]. Journal of health economics, 2015, 39: 147-158.
- [15] Kehler BH, Andersen R, Glaser WA. A Behavioral Model of Families' Use of Health Services[J]. The Journal of Human Resources, 1972, 7(1): 125.
- [16] 蔡伟贤, 吕函梓, 沈小源. 长期护理保险、居民照护选择与代际支持: 基于长护险首批试点城市的政策评估[J]. 经济学动态, 2021(10): 48-63.
- [17] 黄枫, 吴纯杰. 基于转移概率模型的老年人长期护理需求预测分析[J]. 经济研究, 2012, 47(S2): 119-130.
- [18] 朱铭来, 何敏. 长期护理保险会挤出家庭照护吗? ——基于 2011—2018 年 CHARLS 数据的实证分析[J]. 保险研究, 2021(12): 21-38.
- [19] Tamiya N, Noguchi H, Nishi A, et al. Population ageing and well-being: lessons from Japan's long-term care insurance policy[J]. The Lancet, 2011, 378(9797): 1183-1192.
- [20] Kim S, Chae J, Nam SI. Changes in the equity of the long-term care system in Korea based on coulter index differences for the years 2000, 2008, and 2015 [J]. Journal of Aging and Social Policy, 2022, 34(1): 91-107.
- [21] Liu H, Rizzo JA, Fang H. Urban-rural disparities in child nutrition-related health outcomes in China: the role of hukou policy[J]. BMC Public Health, 2015, 15: 1159.
- (责任编辑: 邓 妍)