

中西部农村居民健康状况 Logistic 模型分析*

张永辉, 王征兵

(西北农林科技大学 经济管理学院, 陕西 杨凌 712100)

摘要 运用有序 Logistic 模型分析了我国中西部农村居民的健康状况及其决定因素。研究发现, 近年来我国中西部农村居民健康状况总体有所好转, 中部地区的状况要好于西部地区。中老年人的健康水平下降, 妇女的健康水平较差。中学教育对改善健康水平有积极影响。收入的增加和家庭环境卫生的改善有助于提高健康水平。建议政府继续加大对中西部地区农村合作医疗制度改革的支持力度, 重视对农村基础教育和环境卫生设施的投入, 采取措施维护中老年人和妇女群体的健康水平, 改善落后地区的卫生设施和医疗保健条件, 促进健康公平。

关键词 农村居民; 人力资本; 自评健康状况; 有序 Logistic 模型

中图分类号: C912.82 文献标识码: A 文章编号: 1008-3456(2009)06-0037-05

Logistic Model Analysis on Western and Middle Rural Residents' Health Status and Its Determinants

ZHANG Yong-hui, WANG Zheng-bing

(College of Economics and Management, Northwest A&F University, Yangling, Shaanxi, 712100)

Abstract This paper analyzes the rural residents' health status and its determinants in Western and Middle China. It shows that rural residents' health conditions have been improving, and those in the middle are better than those in the west. Improved education, income and sanitary conditions have positive effect on people's health conditions, while the aged and the female's health conditions are generally worse than the others. It is recommended to strengthen the support on rural cooperative medical scheme in the western and middle rural areas, to give priority to rural education, sanitary as well as environment investment, and to improve the old people and women's health conditions and health consciousness so as to improve health equality.

Key words rural residents; human capital; self-assessed health status; ordered Logistic model

一、问题的提出

健康状况是衡量人力资本质量的重要指标。衡量健康状况通常使用死亡率、患病率、营养状况、伤病影响的未工作天数和自评健康等指标^[1-2]。尽管自评健康指标存在一些不足, 由于它是调查对象身体状况的综合反映, 在研究中经常使用^[3-5]。

对于健康影响因素的研究, 目前主要侧重于个

体特点、社会经济状况和环境卫生条件等三个方面^[6-7]。Grossman^[8]以人力资本理论为基础提出了健康需求模型, 成为研究健康决定因素的基础模型。他认为, 健康是资本品, 个人为了积聚人力资本需要在正规教育、技能培训和医疗保健等方面投资。李和森^[9]认为, 收入增长缓慢以及医疗费用的激增加大了农村居民潜在的健康风险。也有研究指出, 环境污染和卫生条件对健康的影响不可忽视^[10-11]。

收稿日期: 2009-07-08

* 教育部新世纪优秀人才支持计划项目(NCET-05-0859)和西北农林科技大学人才专项基金项目(011404)。

作者简介: 张永辉(1974-), 女, 讲师, 博士研究生; 研究方向: 农村金融和农村社会发展。

我国中西部地区经济发展水平滞后,农村地区社会化服务水平较差,农村居民健康水平令人堪忧。从现有的文献来看,对我国中西部地区农村居民健康状况的研究尚不多见,也没有人对我国农村居民健康水平的时序变化进行过系统的定量研究^[12-14]。本文运用有序 Logistic 模型分析了我国中西部农村居民健康状况的影响因素以及健康水平的区域差异和动态变化规律,以期为农村社会经济的可持续发展和医疗保障制度改革提供政策建议。

二、数据说明和研究方法

1. 数据说明

本项目研究使用中国健康和营养调查(CHNS)数据,该调查涵盖我国不同经济地带 9 个省份城乡居民家庭的生产、生活、收入、消费以及个人营养健康等指标,是比较权威的我国营养健康方面的微观调查数据。考虑研究的目的和成人自我健康评价稳定性较强的特点,本文以我国中西部地区年龄大于 18 周岁的农村居民为研究对象,选择农村合作医疗制度改革前后的 1997 年、2000 年、2004 年和 2006 年的数据进行分析。CHNS 调查的中部省份包括黑龙江、河南、湖北、湖南 4 省,西部省份包括贵州和广西 2 省区(广西是新型农村合作医疗制度的试点地区,享受中央政府对中西部地区省份合作医疗制度的补贴政策,故本文将广西作为西部省份考察)。

2. 研究方法

在社会经济研究中,对于微观主体态度的度量通常使用多分类定性变量,而多分类定性问题的回归分析一般使用有序概率模型。考虑到参数解释的直观性,本文使用有序 Logistic 模型分析,它通常以潜变量的形式出现:

$$Y_k^* = X_k' \beta + \epsilon_k, E[\epsilon_k | X_k] = 0, \epsilon_k \sim (0, \sigma_k^2) \quad (1)$$

这里, Y_k^* 称为潜变量,它是观测不到的取值连续的变量,代表调查对象对自身健康评价的行为倾向。 X_k 是解释变量的向量组,决定着 Y_k^* 的条件均值。 ϵ_k 是随机误差项,服从 Logistic 分布。

有序 Logistic 回归模型中,假设因变量的观测值有 S 种分类($\alpha_1, \alpha_2, \dots, \alpha_s$),潜变量和因变量之间的关系可以表示为:

$$Y = \begin{cases} \alpha_1 & \text{当 } Y^* \leq \tau_1 \\ \alpha_2 & \text{当 } \tau_1 < Y^* \leq \tau_2 \\ \dots & \dots \\ \alpha_s & \text{当 } \tau_{s-1} < Y^* \end{cases} \quad (2)$$

这里, τ_s 是分隔点,共有 $S-1$ 个值。 α_s 表示累计分布函数概率值的临界点, $\alpha_1 < \alpha_2 < \dots < \alpha_s$, 它将 Y^* 的取值依次划分为 S 组。有序 Logistic 模型的数学形式如下:

$$p = P(Y \leq j | X) = 1 - P(Y > j | X) = \frac{e^{X' \beta}}{(1 + e^{X' \beta})} \quad (j = 1, 2, \dots, S-1) \quad (3)$$

这里, X 是解释变量的向量组。 β 是待估参数。在本文中, p 是不同水平的自评健康状况的累积条件概率。Logistic 回归模型的参数是非线性的,通过 Logistic 转换,可以实现模型的线性化,从而使用线性回归方法估计模型。

$$\log\left(\frac{p}{1-p}\right) = \log\left[\frac{P(Y \leq j | X)}{1 - P(Y > j | X)}\right] = X' \beta \quad (j = 1, 2, \dots, S-1) \quad (4)$$

式(4)称为 Logistic 模型。这里, \log 指自然对数。 $p/(1-p)$ 称为发生比(odds),它是事件发生($Y \leq j$)和不发生($Y > j$)的概率之比。这样,发生比的对数形式(Logistic)就表示成了解释变量的线性组合。

实证分析中,为了方便模型系数的解释,通常对式(3)两边取幂,将以 Logistic 为单位的影响作用转换为对应的发生比率(odds ratio),即 e^β 。发生比率说明了解释变量对发生比变动的影响,从而影响因变量发生概率的变动。也就是说,解释变量每增加一个单位,发生比将增加 e^β 倍,因变量发生($Y=1$)的概率也会随之增加。发生比率是大于 0 的正数。当 $\beta > 0$, $e^\beta > 1$, 事件发生的可能性增加;当 $\beta < 0$, $e^\beta < 1$, 事件发生($Y=1$)的可能性减少;当 $\beta = 0$, $e^\beta = 1$, 不论解释变量怎样变化发生比都不会变化。

CHNS 调查中将“自评健康状况”分为“优”、“良”、“中”、“差”四个等级进行考察。对调查对象自评健康的主观倾向无法直接观察,但是可以将他们对自身健康状况的主观判断分类排序,形成四种程度各异的结果,即前文所说的四个不同的临界值。另外,有序 Logistic 模型成立的一个基本假定是发生比成比例,即不同累积对数发生比的回归线相平行,只有截距参数有所差异^[15]。通过构造似然函数,运用最大似然估计方法可以对 Logistic 模型的参数进行估计。

三、实证分析

1. 变量说明和样本分析

本文使用的有效样本共 2 578 个。选取如下指

标作为解释变量建立回归模型(见表 1 和表 2):

(1)年龄。总体来看,样本的平均年龄为 45.87 岁。从时序上看,平均年龄从 1997 年的 41.81 岁增加到 2006 年的 50.03 岁,10 年时间增加了 8 岁多,与我国人口老龄化的发展趋势相吻合。

(2)受教育年限。平均受教育年限为 6.56 年,标准差为 3.07 年。分地区看,西部和中部分别为 6.13 年和 6.77 年。这说明,农村居民文化水平总体偏低,地区之间存在差异。

(3)家庭人口。农村家庭平均人口 4.12 人。值得注意的是,农村家庭规模有缩小的趋势,家庭平均人口从 1997 年的 4.41 人逐渐减少到 2006 年的 3.97 人。

(4)家庭人均收入。按可比价格计算,家庭人均收入逐年增加,从 1997 年的 1 769.94 元增加到 2006 年的 3 262.03 元,年均增幅 6.3%。

表 1 连续型变量的均值(1997—2006 年)

| 变量名称 | 1997 | 2000 | 2004 | 2006 | 1997— 2006 平均 |
|-----------|----------|----------|----------|----------|------------------|
| 平均年龄(岁) | 41.81 | 44.38 | 48.21 | 50.03 | 45.87 |
| 受教育年限(年) | 6.25 | 6.39 | 6.85 | 6.78 | 6.56 |
| 家庭规模(人) | 4.41 | 4.28 | 3.74 | 3.97 | 4.12 |
| 家庭人均收入(元) | 1 769.94 | 1 970.59 | 2 340.15 | 3 262.03 | 2 301.98 |
| 感冒治疗费用(元) | 11.48 | 14.53 | 19.10 | 25.88 | 17.38 |

表 2 离散型变量的样本结构 %

| 变量名称 | 是 | 否 |
|-------|--------------|--------------|
| 婚姻状况 | 2 208(85.65) | 370(14.35) |
| 性别 | 1 058(41.04) | 1 520(58.96) |
| 高血压 | 317(12.30) | 2 261(87.70) |
| 吸烟 | 1 426(55.31) | 1 152(44.69) |
| 饮酒 | 1 815(70.40) | 763(29.60) |
| 有医疗保险 | 868(33.67) | 1 710(66.33) |

说明:性别变量中“是”代表女性;“否”指男性。

(5)感冒治疗费用。平均感冒治疗费用增幅迅速,从 1997 年的 11.48 元增加到 2006 年的 25.88 元。10 年时间增加了 1.26 倍,这反映了近年来医疗费用不断上涨的事实。

(6)婚姻状况和性别。已婚人数为 2 208 人,约占总数的 85.65%;女性为 1 058 人,约占 41.04%。

(7)高血压。高血压患者约占 12.3%,而且,高血压的患病率随年龄增加有明显的增长趋势,在 30 岁以下年龄组中,患病率不到 0.5%。40—49 岁和

50—59 岁年龄组的患病率分别为 4.42%和 8.78%,到了 60 岁以上年龄组,这一指标猛增到 18.95%。

(8)吸烟和饮酒史。分别有 55.31%和 70.4%的被调查者有吸烟和饮酒习惯。

(9)医疗保险。总体上看,有 33.67%的受访者参加了医疗保险,而且人数比例呈逐年上升趋势。从 1997 年的 13.56%增加到 2006 年的 58.87%。这与政府积极推行的新型农村合作医疗制度改革有关。

另外,模型分析中还考察了家庭卫生条件对健康状况的影响,以及中西部农村健康水平的区域差异和时间上的动态变化。

2. 结果分析

基于以上分析,本部分建立有序 Logistic 模型主要考察影响中西部农村居民自评健康状况的人口和社会经济因素。模型 1,模型 2 和模型 3 分别使用西部省份、中部省份和中西部全体样本进行回归分析。模型 4 在模型 3 的基础上添加了年度虚变量(1997 年为参照组)和地区虚变量。使用 STATA 软件分析表明,有序 Logistic 模型的发生比成比例的假设成立。不同模型中, $P < 0.001$,说明模型总体拟合得很好(见表 3)。由于对地区和时间虚拟变量的分析是本项目研究的主要目的之一,加之不同模型之间大多数变量值和显著性水平差别不大,以下主要对模型 4 的结果进行分析。

(1)个体特征因素的影响分析。总体上看,不同模型中各个年龄组的发生比率均小于 1,且呈递减趋势,系数的显著性水平很高。这说明,随着年龄的增加,农村群众的健康状况不断下降,这与生命周期的特征相符。除模型 1 之外,婚姻的发生比率均大于 1,但是系数的显著性不明显。相反,性别的发生比率均小 1,系数均具有统计上的显著性。可以看出,妇女的健康水平比男子差,但是没有证据表明已婚人群的健康状况更好。

从文化水平来看,不同组的发生比率均大于 1,而且有增加的趋势。初中和高中文化水平组的系数均具有统计上的显著性。这说明,教育对健康有积极影响,同文盲人群相比,中学文化水平的健康差异最大。

高血压的系数显著性水平很高,发生比率约为 0.46。也就是说,非高血压患者的发生比率是高血压患者的 2.63 倍(1/0.36)。可以看出,作为一种常见病,高血压对农村人群的身体影响很大。经

表 3 模型结果

| 解释变量 | 模型 1 | | 模型 2 | | 模型 3 | | 模型 4 | |
|---------------------------------|-------------|-----------------|--------------|-----------------|--------------|-----------------|--------------|-----------------|
| | 系数 <i>b</i> | Exp(<i>b</i>) | 系数 <i>b</i> | Exp(<i>b</i>) | 系数 <i>b</i> | Exp(<i>b</i>) | 系数 <i>b</i> | Exp(<i>b</i>) |
| 年龄组(18—29 岁) | 参照组 | | 参照组 | | 参照组 | | 参照组 | |
| 年龄组(30—39 岁) | -0.164 | 0.85 | -0.475 * | 0.62 | -0.287 * | 0.75 | -0.317 * | 0.73 |
| 年龄组(40—49 岁) | -0.526 * | 0.59 | -0.735 *** | 0.48 | -0.547 *** | 0.58 | -0.609 *** | 0.54 |
| 年龄组(50—59 岁) | -0.952 *** | 0.39 | -1.203 *** | 0.30 | -1.055 *** | 0.35 | -1.072 *** | 0.34 |
| 年龄组(60 岁以上) | -1.333 *** | 0.26 | -1.652 *** | 0.19 | -1.466 *** | 0.23 | -1.474 *** | 0.23 |
| 已婚 | -0.104 | 0.90 | 0.089 | 1.09 | 0.041 | 1.04 | 0.034 | 1.04 |
| 女性 | -0.174 * | 0.84 | -0.449 ** | 0.64 | -0.307 ** | 0.74 | -0.356 ** | 0.70 |
| 文化水平(文盲) | 参照组 | | 参照组 | | 参照组 | | 参照组 | |
| 文化水平(小学) | 0.215 | 1.24 | 0.17 | 1.19 | 0.216 | 1.24 | 0.202 | 1.22 |
| 文化水平(初中) | 0.561 ** | 1.75 | 0.523 ** | 1.69 | 0.578 *** | 1.78 | 0.571 *** | 1.77 |
| 文化水平(高中) | 0.567 | 1.76 | 0.548 ** | 1.73 | 0.617 *** | 1.85 | 0.585 *** | 1.79 |
| 文化水平(大学) | -0.21 | 0.81 | 0.671 | 1.96 | 0.098 | 1.10 | 0.236 | 1.27 |
| 高血压 | -0.697 * | 0.50 | -0.803 *** | 0.45 | -0.748 *** | 0.47 | -0.767 *** | 0.46 |
| 吸烟 | 0.262 ** | 1.30 | 0.297 ** | 1.35 | 0.285 ** | 1.33 | 0.277 ** | 1.32 |
| 饮酒 | -0.017 | 0.98 | 0.223 ** | 1.25 | 0.137 ** | 1.15 | 0.128 ** | 1.14 |
| 人均收入(元) | 0.000 | 1.00 | 0.000 * | 1.00 | 0.000 ** | 1.12 | 0.000 ** | 1.16 |
| 家庭人口(人) | 0.096 | 1.10 | -0.081 | 0.92 | -0.018 | 0.98 | -0.022 | 0.98 |
| 环境卫生 | 0.185 * | 1.20 | 0.154 * | 1.17 | 0.166 ** | 1.18 | 0.171 ** | 1.19 |
| 有无医疗保险 | -0.121 | 0.89 | 0.093 | 1.10 | 0.085 | 1.09 | 0.058 | 1.06 |
| 感冒治疗费用(元) | 0.010 | 1.01 | -0.009 | 0.99 | -0.006 | 0.99 | -0.007 | 0.99 |
| 年虚变量(1997 年) | | | | | | | 参照组 | |
| 年虚变量(2000 年) | | | | | | | -0.233 * | 0.79 |
| 年虚变量(2004 年) | | | | | | | -0.203 ** | 0.82 |
| 年虚变量(2006 年) | | | | | | | -0.089 ** | 0.92 |
| 地区虚变量(中部) | | | | | | | 0.407 *** | 1.50 |
| 对数似然比(Log likelihood) | -807.828 6 | | -1 752.696 5 | | -2 592.162 3 | | -2 577.576 1 | |
| 似然比卡方值(LR χ^2) | 133.75 | | 268.85 | | 379.33 | | 408.5 | |
| 调整 R-平方(Pseudo R ²) | 0.070 5 | | 0.071 2 | | 0.068 2 | | 0.078 4 | |

注: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

验表明,烟酒损害人体的健康。但是,模型 4 中吸烟和饮酒的发生比率均大于 1,系数显著性水平都比较高,这同我们的预期相反。一个可能的解释是,吸烟或饮酒者对自己的健康状况比较乐观,这或许是他们保持这些嗜好的原因之一。

收入的显著性水平很高,发生比率大于 1,这说明,收入增加能提高健康水平。很显然,收入水平提高能改善生活条件并增加医疗资源的可及性。

(2)家庭社会因素影响分析。家庭人口的发生比率约为 1,系数统计上不显著,所以没有明显的证据说明家庭规模的变化能够改善成员的健康状况。模型 4 中,家庭环境卫生的发生比率为 1.19,系数的显著性水平高,说明改善居住环境和家庭的卫生条件有助于提高健康水平。

医疗保险的发生比率为 1.06,系数的显著性水平不高。表面上看,农村地区的医疗保险对于居民健康水平影响不大。但是,出现这种现象一方面是因为中西部农村地区过去长期缺乏医疗保险制度,参保人数较少,另外一方面说明,正在开展的新型农村合作医疗制度尚待完善。感冒治疗费用的发生比率约为 1,系数显著性不强。这说明,没有证据显示医疗费用的变化对健康状况的影响作用。

(3)地区和时间因素的影响分析。模型 4 中,年度虚变量的发生比率均小于 1,且呈递增趋势,系数的显著性水平很高。地区虚拟变量系数显著性水平很高,发生比率为 1.5。这说明近年来中西部地区农村居民的健康状况有所改善,而中部地区农村群众的健康状况要比西部好一些。

四、结论和政策建议

本文运用有序 Logistic 模型分析了我国中西部农村居民健康水平的地区差异及其决定因素。研究结果表明,近年来,我国中西部农村居民健康状况总体有所好转,中部地区的健康状况要好于西部地区。老年人的健康水平下降,妇女的健康水平比较差。教育尤其是中学教育对健康有积极影响。高血压人群的健康状况同参照人群的差异明显。收入增加和家庭环境卫生的改善有助于提高健康水平。

以此为基础,本文的政策建议主要有:(1)继续加大对中西部地区新型合作医疗制度改革的支持,不断探索减轻疾病经济负担的途径和方法,提高中西部农村群众对疾病风险的抵御能力。(2)在新农村建设中,加大对中西部地区农村基础教育,提高农

村居民的文化水平,采取积极措施改善农村地区的卫生环境和生活条件。(3)结合我国人口老龄化的国情,探索提高农村地区中老年人群医疗保健和特殊护理服务水平的措施;充分利用医疗服务体系和大众传媒手段积极宣传健康生活的理念,提高农村居民尤其是妇女的健康保健意识。(4)高度重视健康水平的地区差异,积极改善落后地区的卫生设施和医疗保健条件,促进健康公平。

参 考 文 献

- [1] WAGSTAFF A, PACI P, DOORSLAER E. On the measurement of inequality in health[J]. Soc Sci Med, 1991, 33(5): 545-557.
- [2] ALLISON R A, FOSTER J E. Measuring health inequality using qualitative data[J]. Journal of Health Economics, 2004, 23(2): 505-524.
- [3] DOORSLAER E, GERDTHAM U G. Does inequality in self-assessed health predict inequality in survey by income? [J]. Social Science and Medicine, 2003, 57(9): 1621-1629.
- [4] 赵忠, 侯振刚. 我国城镇居民的健康需求与 Grossman 模型[J]. 经济研究, 2005(10): 79-90.
- [5] 谷琳, 乔晓春. 我国老年人健康自评影响因素分析[J]. 人口学刊, 2006(6): 25-29.
- [6] BOLLEN K A, GLANVILLE J L, STECKLOV G. Socioeconomic status and class in study of health in developing countries[J]. Annual Review of Sociology, 2001(27): 153-185.
- [7] 王曲, 刘民权. 健康的价值及若干决定因素: 文献综述[J]. 经济学季刊, 2005(10): 1-35.
- [8] GROSSMAN M. On the concept of health capital and the demand for health[J]. Journal of Political Economy, 1972, 80(2): 223-255.
- [9] 李和森. 中国农村医疗保障制度研究[M]. 北京: 经济科学出版社, 2005: 168.
- [10] CROPPER M L. The health benefit of air pollution in New Delhi[J]. American Journal of Agriculture Economics, 1997, 79(5): 1625-1629.
- [11] 韩俊, 罗丹. 中国农村卫生调查[M]. 上海: 上海远东出版社, 2007: 103.
- [12] 赵忠. 我国农村人口的健康状况及影响因素[J]. 管理世界, 2006(3): 78-85.
- [13] 张永辉, 王征兵. 西部农村家庭医疗保健消费影响的实证分析[J]. 西北工业大学学报: 社会科学版, 2008(9): 54-57.
- [14] 申志伟, 蒋远胜. 西部农村居民健康及其家庭医疗支出的决定因素[J]. 农业技术经济, 2008(3): 58-64.
- [15] 王济川, 郭志刚. Logistic 回归模型: 方法和应用[M]. 北京: 高等教育出版社, 2001: 124.

(责任编辑: 刘少雷)