Statistics and Application 统计学与应用, 2018, 7(4), 438-444
Published Online August 2018 in Hans. http://www.hanspub.org/journal/sa
https://doi.org/10.12677/sa.2018.74050

Hans汉斯

Multi-Level Model Analysis of Factors Affecting Chinese Residents' Income

Yinmeng Lin

School of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming Yunnan
Email: lylinyimeng@126.com

Received: Aug. 3rd, 2018; accepted: Aug. 20th, 2018; published: Aug. 27th, 2018

Abstract

Based on the factors influencing the income of residents in China's Comprehensive Social Survey (CGSS) 2013 survey data, a multi-level model analysis of Chinese residents' personal income is conducted on the basis of provinces/regions/municipalities. First of all, this paper introduces the application background and theoretical basis of the multi-level model. Secondly, it introduces the cleaning ideas and variable settings of the data. In the last part, the process of multi-level modeling and the summary analysis of the factors affecting income are discussed in detail. This article considers China's special national conditions: the popularity of education, the general employment of Chinese women, the difference between urban and rural income, and on this basis, it explores gender, ethnicity, marital status, urban or rural areas, parental education, political outlook and education. The impact of years and various cross-cutting factors on income, combined with China's development status and China's social situation, gives a reasonable explanation for the reasons for the influence of variables on personal income in the process of multi-level model exploration.

Keywords

Income Influencing Factors, Political Effects, Educational Returns, Multi-level Models

中国居民收入影响因素多水平模型分析

林义孟

云南财经大学统计与数学学院，云南 昆明
Email: lylinyimeng@126.com

收稿日期：2018年8月3日；录用日期：2018年8月20日；发布日期：2018年8月27日

摘 要

基于中国综合社会调查(CGSS) 2013年度调查数据针对居民收入的影响因素，在省份、地区、直辖市的

文章引用: 林义孟. 中国居民收入影响因素多水平模型分析[J]. 统计学与应用, 2018, 7(4): 438-444.
DOI: 10.12677/sa.2018.74050

基础上对中国居民个人收入进行多水平模型分析。首先，本文介绍了多水平模型的应用背景，以及理论基础；其次，介绍了数据的清洗思想和变量设置，在最后一部分详细论述了多水平建模的过程,以及对收入影响因素的总结分析。本文考虑到中国的特殊国情：教育的普及，中国女性的普遍就业，城市与农村收入差别，并在此基础上探讨了性别、民族、婚姻状态、城市或者农村、父母教育程度、政治面貌，教育年限及各种交叉因素对收入的影响，结合中国发展状态和中国的社会情况给出多水平模型探索过程中变量对个人收入的影响原因的合理解释。

关键词

收入影响因素，政治效应，教育回报，多水平模型

1. 文献综述及背景分析

收入问题受到每一个个体的关心，而每个国家有每个国家的国情，每个省份有每个省份的发展程度和发展速度，忽略不同地区的个体情况，把中国看作一个同质整体，个体放在中国的大环境中分析个人的收入影响因素有一定的合理性但又不是最恰当的分析方法，因为这样忽略了当代中国省份，城市收入不平等的时间和空间变异。中国各地区的经济活动，教育程度在很大程度上是由各地的自然资源、政府政策和人力资源等因素决定的，更重要的是，中国的工业经济改革是分地区进行的，对教育的支持程度也是不同的，对应个体普遍的受教育程度是有分层的。从上述背景出发，本文分地区探讨与个人收入相关因素的关系。

此前，谢宇和韩怡梅(Xie & Hannum, 1996)采用四 88 年中国居民收入调查(CHIP)数据中城市居民的调查数据，以各地区改革步伐的不均衡为前提，研究了经济改革的成功与个人收入决定因素之间的关系[1]。Hauser 和 Xie (2005)发现在经济改革快速推进时期，党员的好处不但依然存在，而且不论被访者受教育程度的高低、工作经历的长短、是男性还是女性，党员的相对优势都有所扩大[2]。对此，一个可能的解释是党员身份也许可以被看作是一种能力，它代表了人力资本中未被观测到的方面(Gerber, 2000; 2001) [3]。因而，党员身份回报的明显提高反映的可能是对某些能力回报的提高。Hauser 和 Xie (2005)揭示了 1998、1995 年间收入决定因素重要性的诸多变化：教育的收入回报对男女两性都有显著提高；收入的性别差异在扩大；党员身份的收入回报也翻了番。也就是说，在个体层次上，收入决定因素的确存在统计上的显著变化[3]。

2. 多水平模型理论

2.1. 多水平模型介绍

人与人之间是有差异的，而且总是生活在一定的社会环境中，其表现和行为方式总是随着其置身于其中的社会环境的变化而变化。在更一般的意义上，社会研究经常会涉及个体与社会环境之间的互动关系。个体会受到其所属的团体氛围或社会环境的影响；反过来，团体氛围或社会环境的属性也会受到作为它们构成要素的个体的影响。个体与社会环境之间的这种互动关系决定了社会研究所用数据中的多层结构。

多水平模型，社会学研究者称其为多层线性模型(multilevel models)，教育学研究者将其称为分层线性模型(hierarchical linear models)，计量经济学者往往称其为随机系数回归模型(random-coefficient regression models)，统计学家则更多地称其为混合效应模型(mixed effect models)和随机效应模型(random-effects models)，而发展心理学研究者多称其为增长曲线模型(growth-curve models)。另外，多层线性模型有一个特例，被习惯称作协方差成分模型(covariance components models)。尽管名称繁多，但多层线性模型大体上包括两个方面的来源：情景分析和混合效应模型[4]。

2.2. 多水平模型基本原理

多水平模型的分析思路其实比较简单。它首先将多层结构数据在因变量上的总变异明确区分成组内和组间两个层次，然后分别在不同的层次上引人自变量来对组内变异和组间变异加以解释。最简单的多层线性模型由一个组内方程和一个组间方程构成，同时将组内方程的部分或全部参数作为结果变量由组间方程来加以解释。

对于多层结构数据而言，变量的变异同样可以区分为组内变异和组间变异两个部分。如果完全忽略组间变异的话，残差分布有可能出现异方差，采用常规最小二乘法所得到的参数估计值尽管仍是无偏和一致的，但不再是最有效的。多水平模型通常使用最大似然估计(简称 MLE)方法来估计模型的方差协方差。但是，在具体应用中，最大似然估计方法又分成完全最大似然法(简称 FML)和限制性最大似然法(简称 REML)两种。两者之间的差别在于它们对模型残差项的考虑有所不同。REML 包含了所有来源的残差，而且它常被用于估计高层次上的单位数量偏少的模型，而在进行模型比较时通常采用 FML 方法进行模型估计[5]。多水平模型子模型[6]主要有空模型又叫截距模型、随机系数模型、完全模型。

3. 多水平模型数据说明[7]

3.1. 数据来源

数据来源于中国综合社会调查(CGSS) 2013 年数据，此数据于 2015 年 1 月 1 日在中国国家调查数据库网站正式发布。中国综合社会调查(Chinese General Social Survey, CGSS)始于 2003 年，是我国最早的全国性、综合性、连续性学术调查项目。CGSS 系统全面的收集社会、社区、家庭、个人多个层次的数据，总结社会变迁的趋势，探讨具有重大科学和现实意义的议题。此调查项目关键字：价值观，健康，截面调查，家庭，教育，职业；地理覆盖范围：中国大陆；分析单位：个人，家庭；人口覆盖范围：18+；调查的时间组织方式：过续性截面调查；调查频率：年度；抽样程序：多阶分层抽样；调查模式：面访。

3.2. 数据分布

2013 年为中国综合社会调查(CGSS)第二期(2010~2019)的第 4 次年度调查，也是 CGSS 自 2003 年开始以来的第 10 年。调查在全国一共抽取了 100 个县(区)，加上北京、上海、天津、广州、深圳 5 个大城市，作为初级抽样单元。在全国一共调查 480 个村/居委会，每个村/居委会调查 25 个家庭，每个家庭随机调查 1 人，总样本量约为 12,000。分布情况见图 1。

3.3. 数据解释说明

原数据变量数：722 个，样本量：11,438 例；经过年龄，空缺，异常值等条件数据清洗收集关于收入和教育因素相关的变量，进入样本 4949 例。表 1 为中国综合社会调查问卷中相关变量及本论文个人派生变量，并对部分变量进行建模分析。

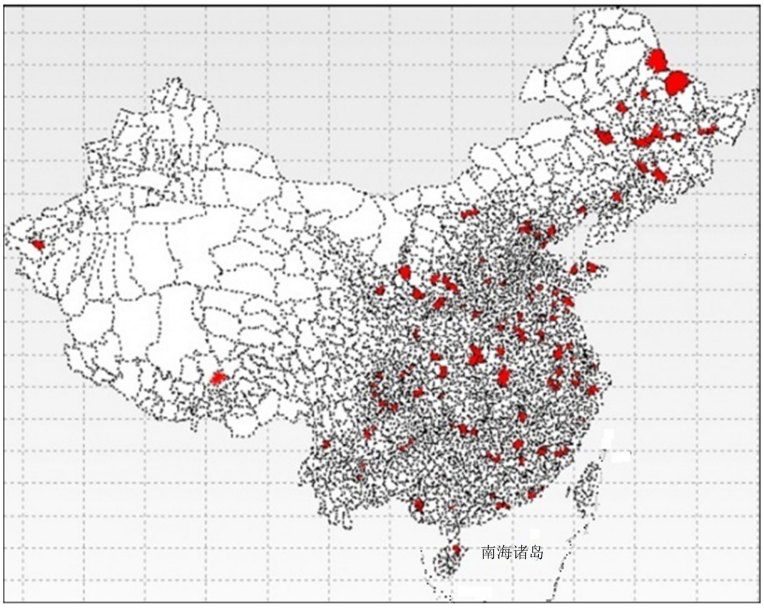

Figure 1. The geographical distribution of the data (source: CGSS)
图 1. 数据地域分布情况(来源：CGSS)

Table 1. Variable description in the questionnaire
表 1. 调查问卷中变量说明

变量名称	变量解释
id	问卷编号
vilorngh	是村委会还是居委会，0 表示居委会，1 表示村委会
s41	采访地点——省/直辖市/自治区
s42	采访地点——市+区/县编码
s43	采访地点——乡镇/街道编码
s5a	受访者居住的地区类型是
a2	性别，0 表示男性，1 表示女性;
a3a	出生日期——年
a4	民族
a4a	民族是——其他
a7a	最高教育程度是(包括目前在读的)
a8a	个人去年(2012)全年的总收入是多少
a10	政治面貌，1 表示中共党员或民主党派，0 表示共青团员、表示群众，用来作为对政治资本的测量
a58	工作经历及状况是什么
a62	去年(2012)全年家庭总收入是多少
a63	目前住在一起的通常有几人(包括您本人)
a69	目前的婚姻状况，1 表示初婚有配偶，0 表示非初婚有配偶
a89b	父亲的最高教育程度是
a90b	母亲的最高教育程度是
age	年龄(个人派生)
lnself_y	个人收入对数(个人派生)
lnfamily_y	家庭收入对数(个人派生)

4. 中国居民收入影响因素分析

4.1. 变量选择说明

从居民收入的实际情况出发，对中国居民收入影响因素分析这一研究选择如下变量：

lnself_y：表示收入的对数，收入指的是个体每年所得的各种来源的收入之和；

vilorngh：0 表示居委会，1 表示村委会；

a2 (性别)：0 表示男性，1 表示女性；

a7a (教育年限)：正整数表示所受教育的时间；

a10 (政治面貌)：1 表示中共党员或民主党派，0 表示共青团员、表示群众，用来作为对政治资本的测量；

a69：1 表示初婚有配偶，0 表示非初婚有配偶；

a2* a7a：教育与性别的交互项，即允许教育回报因性别的不同而不同；

a7a* a10：教育与政治面貌的交互项，即允许教育回报因政治面貌的不同而不同；

a2* a10：性别与政治面貌的交互项。对如上变量进行多水平模型分析如下：

4.2. 空模型检验

以省份\直辖市\自治区为划分，个人收入对数为因变量进行空模型检验，表 2 为空模型分析结果。

P 值均不超过 0.005，省份地区差异的估计值为 0.25031，不同省份地区收入水平有差异，且由公式(1~4)，求得 ICC 为 0.24，从数据来看总方差有 24%的方差比例是由组间方差所影响的，分省地区考虑收入水平具有统计学意义，应进一步拟合多水平模型。

4.3. 全模型分析

如表 3 所示，在首次探索中，以省份\直辖市\自治区为主题，个人收入取对数为因变量，在固定效应中加入 a2 (性别)，a4 (民族)，a7a (教育年限)，a10 (政治面貌)，a69 (婚姻状况)，a89b (父亲教育年限)，a90b (母亲教育年限)，age (2012 年年龄)，vilorngh (城市或农村)，a7a*a10 (教育年限与政治面貌交互项)，a2*a7a (性别与教育年限交互项)，a2*a10 (性别与政治面貌交互项)；在随机效应中加入 vilorngh (城市或农村)，发现在固定效用中民族，政治面貌、父母教育程度，a2*a10 对个人的收入影响不显著则在下一轮模型构建中剔除此类变量。

对于民族(a4)不显著，一方面是因为中国少数民族不管是占中国总人口还是调查人数所占比例均较少，统计的样本太少(少数民族占样本的 8.1%，且分为七大部分：蒙，满，回，藏，壮，维，其他)从而各部分没有到达一定的显著性，二是因为中国近年来的各地区经济发展速度较快加之政府对少数民族地区的扶持力度，各民族共同发展共同繁荣，民族因素在收入差异上没有十分明显的不同。

对于政治面貌、政治面貌和教育年限的交叉影响，不显著，这与此前研究中共党员身份在中国是一种政治优势的理论相反[6]。性别与政治面貌(a2*a10)交叉项的不显著，在样本中男女比例接近一比一，样本中共党员占比超过 10%，大致和人口特点相符，所以不存在样本数量的层面上的影响，这里正是中国男女平等，社会公平方面的进步印证。

如表 4 所示，在第二次探索中，以省份\直辖市\自治区为主题，个人收入取对数为因变量，在固定效应中加入 a2，a7a，a69，age，vilorngh，lnfamily_y，a2*a7a，其中 a2，a69，vilorngh 即：性别、婚姻状态、城市农村为哑变量；在随机效应中加入 vilorngh，各变量均显著。

此时全模型为：

Table 2. Estimates of covariance parameters
表 2. 空模型运行结果

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Residual		0.800161	0.016131	49.604	0.000	0.769162	0.832410
Intercept [subject = s41]	Variance	0.250310	0.069433	3.605	0.000	0.145334	0.431113

a. Dependent Variable：个人收入对数。

Table 3. Complete model fixed effect parameter estimation
表 3. 完全模型固定效应参数估计

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	1.210398	0.132624	2955.107	9.127	0.000	0.950354	1.470442
[vilorngh=0]	0.221933	0.035115	23.039	6.320	0.000	0.149298	0.294568
[vilorngh=1]	0[b]	0	.	.	.	.	.
a7a	0.050287	0.004348	4940.962	11.565	0.000	0.041763	0.058811
[a69=0]	0.062774	0.022950	4932.519	2.735	0.006	0.017782	0.107765
[a69=1]	0[b]	0	.	.	.	.	.
age	−0.003098	0.000949	4939.235	−3.264	0.001	−0.004959	−0.001237
lnfamily_y	0.756516	0.011486	4647.095	65.862	0.000	0.733998	0.779035
[a2=0] * a7a	−0.044902	0.005230	4911.881	−8.586	0.000	−0.055154	−0.034650
[a2=1] * a7a	0[b]	0	.	.	.	.	.
[a2=0]	0.750235	0.033748	4920.081	22.230	0.000	0.684073	0.816397
[a2=1]	0[b]	0	.	.	.	.	.

a. Dependent Variable：个人收入对数。
b. This parameter is set to zero because it is redundant.

Table 4. Complete model random effect parameter estimation
表 4. 完全模型随机效应参数估计

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Residual		0.313758	0.006346	49.441	0.000	0.301564	0.326446
Intercept [subject = s41]	Variance	0.006867	0.004943	1.389	0.165	0.001675	0.028149
vilorngh [subject = s41]	Variance	0.009675	0.004639	2.086	0.037	0.003780	0.024764

a. Dependent Variable：个人收入对数。

层 1 模型：
$$\ln self\_y_{ij} = \beta_{0j} + \beta_{1j} vilorngh_{ij} + \alpha_1 a7a_{ij} + \alpha_2 a69_{ij} + \alpha_3 age_{ij} + \alpha_4 a69_{ij} + \alpha_5 \ln family\_y_{ij} + \alpha_6 a2_{ij} * a7a_{ij} + \alpha_7 a2_{ij} + e_{ij}$$

层 2 模型：
$$\beta_{0j} = \gamma_{00} + \gamma_{01} vilorngh_j + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + u_{1j}$$

组合模型：
$$\ln self\_y_{ij} = \gamma_{00} + \gamma_{01} vilorngh_{ij} + \alpha_1 a7a_{ij} + \alpha_2 a69_{ij} + \alpha_3 age_{ij} + \alpha_4 a69_{ij}$$
$$+\alpha_5 \ln family\_y_{ij} + \alpha_6 a2_{ij} * a7a_{ij} + \alpha_7 a2_{ij} + \left(u_{0j} + u_{1j} vilorngh_{ij} + e_{ij}\right)$$

对固定效应进行检验，均有 $P < 0.05$，说明教育年限、婚姻状态、年龄、城市或农村、家庭收入水平，性别，性别与教育年限交互项均对个人收入都有影响。

5. 中国居民收入影响因素多水平模型分析结论

婚姻状态为初婚有配偶的相对于非初婚有配偶的系数 0.062774，一方面说明婚姻状态对事业有正影响，一方面暗示婚姻状态也是其他能力的一种体现。

城市的对收入影响的系数为正，对收入有正的影响，体现了城市的就业机会、工资水平、发展可能性比农村要有优势[3] [8]。

年龄对收入影响的系数为−0.003098，系数为负且较小，说明在青壮年时期年龄因素对收入的影响效应不大，但是随着年龄继续升高，对收入的累计效应将有负的影响，符合现在的社会现状。

家庭收入对个人收入的影响系数为 0.756516，这似乎在说明一种类似“书香门第”的传递效应，也可能是因为家庭成员收入相当，总体水平的上涨是因为个体水平的积累效应。

男性相对于女性的收入，0.750235 说明男性工资水平比女生高。在性别与教育年限(a2*a7a)较差效应为负，男性相对于女性的教育回报率为负，系数为−0.44902，说明女性的教育回报率要高于男性的教育回报率，这可能反映了受教育程度最少的那部分工作者而言，男性和女性可获得的工作类型有所不同，收入的性别差异在继续扩大，这支持了有关经济改革可能加具劳动力市场性别不平等的结论[9]。这可能是由于收入的性别差距在低受教育水平上表现很大，但是随着受教育水平的上升，收入的性别差异在缩小，甚至不明显，这与文献中探究的结果一致[10] [11]。

参考文献

[1] Xie, Y. and Hannum, E. (1996) Regional Variation in Earning Inequality in Reform-Era China. *American Journal of Sociology*, **101**, 950-992. https://doi.org/10.1086/230785

[2] Hauser, S.M. and Xie, Y. (2005) Temporal and Regional Variation in Earnings Inequality:Urban China in Transition Between1988 and 1995. *Social Science Research*, **34**, 44-79. https://doi.org/10.1016/j.ssresearch.2003.12.002

[3] Gerber, T.P. (2000) Membership Benefits or Selection Effects? Why Former Communist Party Members Do Better in Post-Soviet Russia. *Social Science Research*, **29**, 25-50. https://doi.org/10.1006/ssre.1999.0651

[4] 谢宇. 回归分析[M]. 北京: 社会科学文献出版社, 2010: 294-317.

[5] 王济川, 谢海义, 姜宝法. 多层统计分析模型——方法与应用[M]. 北京: 高等教育出版社, 2008: 78-79.

[6] 石磊, 向其凤, 陈飞. 多水平模型及其在经济领域中的应用[M]. 北京: 科学出版社, 2013: 49-82

[7] 中国综合社会调查(CGSS)中国国家调查数据库. http://cnsda.ruc.edu.cn/index.php?r=projects/view&id=93281139

[8] 韩东林, 付鹏. 人力资本投入差异与城乡居民收入差距——基于省级面板数据的实证分析[J]. 技术经济与管理研究, 2014(6): 28-32.

[9] 高云, 詹慧龙, 陈伟忠, 矫健. 我国农村居民收入现状与影响因素分析[J]. 江西农业大学学报(社会科学版), 2013, 12(2): 178-185.

[10] Shu, X.L. and Bian, Y.J. (2002) Intercity Variation in Gender Inequalities in China: Analysis of a 1995 National Survey. *Research in Social Stratification and Mobility*, **19**, 267 -307. https://doi.org/10.1016/S0276-5624(02)80044-0

[11] 虢超, 丁建军. “关系”和教育对中国居民收入的影响——基于 CGSS 调查数据的实证分析[J]. 南方经济, 2014(3): 38-51.

知网检索的两种方式：

1. 打开知网页面 http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD
   下拉列表框选择：[ISSN]，输入期刊 ISSN：2325-2251，即可查询
2. 打开知网首页 http://cnki.net/
   左侧“国际文献总库”进入，输入文章标题，即可查询

投稿请点击：http://www.hanspub.org/Submission.aspx
期刊邮箱：sa@hanspub.org