

教育对居民个人收入的影响分析

——基于2017年中国综合社会调查数据

靳雨西

上海工程技术大学, 上海

收稿日期: 2022年6月19日; 录用日期: 2022年7月18日; 发布日期: 2022年7月25日

摘要

教育是提高人民综合素质、促进人的全面发展的重要途径, 是民族振兴、社会进步的重要基石, 是对中华民族伟大复兴具有决定性意义的事业, 更是改善家庭环境、提高人民生活水平的重要渠道。基于2017年中国综合社会调查数据, 利用多元回归的嵌套式模型分析教育对居民收入的影响, 研究得出, 居民的受教育水平、配偶的受教育水平、父亲母亲的受教育水平对居民收入有正向的影响, 这就要求我们加强对教育的培养, 重视教育的发展。

关键词

教育水平, 收入, 回归分析

Analysis of the Impact of Education on the Personal Income of Residents

—Based on the Data of the 2017 China Comprehensive Social Survey

Yuxi Jin

Shanghai University of Engineering Science, Shanghai

Received: Jun. 19th, 2022; accepted: Jul. 18th, 2022; published: Jul. 25th, 2022

Abstract

Education is an important way to improve the people's comprehensive quality and promote the all-round development of the people, an important cornerstone of national rejuvenation and social progress, a cause of decisive significance to the great rejuvenation of the Chinese nation, and an important channel for improving the family environment and raising the people's living stan-

dards. Based on the data of the 2017 China Comprehensive Social Survey, the nested model of multiple regression is used to analyze the impact of education on residents' income, and the study shows that the education level of residents, the education level of spouses, and the education level of fathers and mothers have a positive impact on residents' income, which requires us to strengthen the cultivation of education and attach importance to the development of education.

Keywords

Educational Level, Revenue, Regression analysis

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

居民收入水平的影响因素研究是经济学和社会学长期关注的重要问题之一，对促进社会公平和经济社会发展具有重要意义，在经济学中，著名的明瑟方程就研究了这一重要问题。改革开放以来，居民的生活水平大幅提高，人们一般认为人力资本的投入可以明显提升居民个人生活水平，尤其是教育的投入，教育对收入的影响具有很大的现实意义。客观认识这一问题可以使人们提升人们对教育的重视程度，从而有助于人们做出更为理性的决策和合理的规划，同时教育具有较强的政策意义，收入的增长是较少贫困的重要组成部分，认识到教育的政策意义有助于增加居民收入，减少贫困，促进社会公平。

关于教育对收入的影响，许多研究者对其进行探讨分析。学者董良采用明瑟方程和“布劳-邓肯”模型研究了教育对个人收入的影响，经过深入的探讨，发现接受教育水平越高，收入越高[1]；谢周亮使用 2005 年中国综合社会调查数据，采用基于回归方程的因素分解法，对造成个人收入差异的因素及其影响程度进行比较分析，结果显示，个人教育回报随着家庭教育背景的提高而提高，并且母亲的教育背景对个人收入的正面影响相对较大[2]；曹麦基于扩展的明瑟收入方程，采用 2013 年中国综合社会调查数据分析学校教育和家庭教育对个体总收入的影响，结果显示，学校教育和家庭教育均对个体收入有显著的正向作用[3]；徐秋艳和张秋炜利用中国家庭追踪调查 2012 年和 2016 年的数据，讨论了家庭受教育水平和健康水平与收入水平的关系，他们发现，受教育水平高的家庭更有可能拥有更高的收入[4]；张津琛、储宁、孙东宁利用 2015 年中国综合社会调查数据指出教育对居民个人收入具有正向的影响，将教育看作是一种预测人未来生产率的信号，个人的受教育年限越长，收入水平也就越高[5]；李秀玉、蔡玉洁利用 2015 年中国综合社会调查数据强调了质量可比的教育对居民个人的收入具有正向的影响[6]。

综上所述，大多数研究者都强调了教育对居民收入的影响，他们可能更偏向于从个人的受教育年限的角度进行分析，本文则试图从教育的多个方面进行剖析，如个人的受教育程度、配偶或同居伴侣的受教育程度、父亲的受教育程度、母亲的受教育程度、教育和培训支出等方面，利用 2017 年中国综合社会调查数据分析教育对居民收入的影响。

2. 数据来源与统计方法

2.1. 数据来源

中国综合社会调查是在全国范围内进行调查，样本量很大，数以万计，是我国最早的全国性、综合

性社会调查,是研究中国社会最主要的数据来源,非常具有权威性。中国人民大学2017年进行的“中国综合社会调查”是一项全国范围内的大型随机抽样调查,调查问卷数量多且问卷内容充实,利用该调查数据进行研究得出的结论可以推论全国,提供了大量详实且微观的数据。利用stata对数据进行处理之后,关于年收入的有效样本数据达到了11,900,基本上可以代表中国居民的真实情况。

2.2. 统计方法

本文采用统计分析软件stata16.0进行分析,对居民受教育程度、配偶或同居伴侣受教育程度、父亲的受教育程度、母亲的受教育程度和教育培训支出进行嵌套式多元回归分析,分析这些变量对因变量居民收入的影响。

2.3. 变量说明与描述

本文从教育入手分析对居民收入的影响,并且从教育的多个方面进行研究,因此选取居民的受教育程度、配偶或同居伴侣的受教育程度、父亲的受教育程度、母亲的受教育程度以及教育培训支出作为自变量,居民年收入作为因变量,由于本文研究的是教育对收入的影响,将年龄、性别、家庭收入等作为控制变量。因为居民的受教育程度、配偶或同居伴侣的受教育程度、父亲的受教育程度、母亲的受教育程度四个变量属于类别变量,因此采取tab的命令对其进行描述,教育培训支出和居民年收入属于连续变量,因此采取sum的命令对其进行描述。

从表1可以看出,居民2016年的收入平均水平达到了4万元,标准差显示大部分居民年收入值与均值具有较大的差异,说明居民收入差距相对来说较大,通过表格中的最大值和最小值也可以显示出来,这也从侧面说明居民的年收入的分配不均匀,需要对年收入值取自然对数;第二行的教育培训支出的均值到达了27,000元,标准差较小,说明教育培训支出与平均值相比差距较小,但是从最大值最小值来看,居民教育培训支出的差距还是比较大的。

Table 1. Variable description

表 1. 变量描述

Variable	Obs	Mean	Std. Dev.	Min	Max
yincom	11,900	41913.16	250567.5	0	9,999,996
educmoney	3789	27089.26	3.29168	1	9,999,997

从表2可以看出,最高受教育程度一栏的样本量达到了12,582人,样本数据充足,并且可以看出目前在读高中以下年级的人占大多数,所占比例高达62.44%,其中目前受教育程度是初中的人数最多,达到了3500人,占比27.9%;高中阶段受教育程度是普通高中的人数最多,有1470人,占比11.68%,并且和读其他高中相比差距较大;大学以上阶段大约占比20%,相对来说较为均衡,但是在大学阶段,最高受教育程度是大学本科的人数最多,占比最大。

从配偶或同居伴侣最高受教育程度这一栏可以看出,大部分人的学历集中于高中以下阶段,占样本总数的65.7%,但是其中大部分人数还是接受过小学和初中的教育,没有受过教育的人占比较小,接受过私塾和扫盲班教育的人数少之又少;高中阶段人们的受教育情况则主要集中于普通高中,接受普通高中教育人数占比12.06%;大学阶段接受普通本科教育的人数最多,占比5.8%。

从父亲最高教育程度一栏可以看出大部分人学历集中于高中以下阶段,高中学历一下人群占样本比

例达到了 85.96%，其中没有接受过教育的人数最多，占比最大；高中阶段接受普通高中教育的人数最多，占比较大；大学阶段接受大学本科教育的人数相对来说较多，占比 1.95。

从母亲最高教育程度一栏中可以看出绝大部分人接受过高中以下的教育，高中以下学历人数达到了 91.98，其中 58.45% 的人没有接受过教育，是接受过私塾、小学初中教育人数的一半；高中和大学阶段大部分人数接受过普通高中和本科教育。

总体来看，大多数人接受过比较系统的教育，经历了从小学、初中、普通高中和大学本科的教育，但是相对于父亲母亲接受的教育程度来说，居民自己和配偶的教育程度相对较高，父亲母亲大部分人甚至都没有接受过教育，这也从侧面反映了随着社会的发展，人们也越来越注重教育的作用，追求高学历的人群越来越多，没有接受过教育的人越来越少。

Table 2. Describes the statistical results

表 2. 描述统计结果

	您目前的最高教育程度			配偶或同伴最高受教育程度			父亲最高教育程度			母亲最高教育程度		
	Freq.	Percent	Cum.	Freq.	Percent	Cum.	Freq.	Percent	Cum.	Freq.	Percent	Cum.
没有受过任何教育	1528	12.14	12.14	1097	11.33	11.33	4570	38.89	38.89	7058	58.45	58.45
私塾、扫盲班	91	0.72	12.87	40	0.41	11.75	897	7.63	46.53	392	3.25	61.7
小学	2726	21.67	34.53	2255	23.3	35.05	2837	24.14	70.67	2426	20.09	81.79
初中	3511	27.9	62.44	2966	30.65	65.7	1796	15.29	85.96	1230	10.19	91.98
职业高中	163	1.3	63.73	121	1.25	66.95	59	0.50	86.46	33	0.27	92.25
普通高中	1470	11.68	75.42	1167	12.06	79	833	7.09	93.55	484	4.01	96.26
中专	530	4.21	79.63	422	4.36	83.36	190	1.62	95.17	161	1.33	97.59
技校	72	0.57	80.2	60	0.62	83.98	23	0.20	95.36	20	0.17	97.76
大学专科(成人高等教育)	377	3	83.2	231	2.39	86.37	64	0.54	95.91	42	0.35	98.10
大学专科(正规高等教育)	674	5.36	88.56	441	4.56	90.93	121	1.03	96.94	76	0.63	98.73
大学本科(成人高等教育)	299	2.38	90.93	191	1.97	92.9	49	0.42	97.35	29	0.24	98.97
大学本科(正规高等教育)	948	7.53	98.47	561	5.8	98.7	229	1.95	99.30	104	0.86	99.83
研究生及以上	172	1.37	99.83	114	1.18	99.88	15	0.13	99.43	4	0.03	99.87
其他	21	0.17	100	12	0.12	100	67	0.57	100	16	0.13	100
总计	12,582	100		9678	100		11,750	100		12,075	100	

3. 模型分析

3.1. 模型解释

本文探讨分析教育对各方面对居民收入的影响，选取居民目前的受教育程度、配偶或同居伴侣的受教育程度、父亲的受教育程度、母亲的受教育程度以及教育培训支出作为自变量，居民年收入作为因变量，分析多个自变量对一个因变量的影响，因此选取多元回归中的嵌套模型，根据变量构建以下四个模型：

$$\text{模型 1: } y(\text{yincom}) = a + b_1 * x_1(\text{edulevel})$$

$$\text{模型 2: } y(\text{yincom}) = a + b_1 * x_1(\text{edulevel}) + b_2 * x_2(\text{peduclevel})$$

$$\text{模型 3: } y(\text{yincom}) = a + b_1 * x_1(\text{edulevel}) + b_2 * x_2(\text{peduclevel}) + b_3 * x_3(\text{feduclevel})$$

$$\text{模型 4: } y(\text{yincom}) = a + b_1 * x_1(\text{edulevel}) + b_2 * x_2(\text{peduclevel}) + b_3 * x_3(\text{feduclevel}) + b_4 * x_4(\text{meduclevel})$$

$$\text{模型 5: } y(\text{yincom}) = a + b_1 * x_1(\text{edulevel}) + b_2 * x_2(\text{peduclevel}) + b_3 * x_3(\text{feduclevel}) + b_4 * x_4(\text{meduclevel}) + b_5 * x_5(\text{educmoney})$$

“yincom”为因变量居民年收入，“edulevel”为居民目前的最高受教育程度，“peduclevel”表示配偶或同居伴侣的最高受教育程度，“feduclevel”表示父亲的最高教育程度，“meduclevel”表示母亲的最高教育程度，“educmoney”表示教育培训支出。

3.2. 模型分析

3.2.1. 模型 1 结果分析

在模型 1 中，居民年收入为因变量，居民的最高受教育程度为自变量。此模型为基准模型。

从图 1 中可以看出，模型 1 是显著的， p 值 = $0.0000 < 0.05$ ，R-squared 是 0.2808，表明其解释力度为 0.2808，这个模型解释了 28.08% 的收入受到居民受到的最高教育程度的影响，自变量居民的最高受教育程度的 p 值 < 0.05 ，说明自变量对因变量的影响是显著的。根据上图可以得出模型 1 的公式，即 $y = 8.799824 + 0.2157 * x_1$ ，也就意味着居民的受教育程度上升一个水平，居民收入会增加 0.2156922 元。

```
. reg logincom edulevel
```

Source	SS	df	MS	Number of obs =	9,991
Model	4992.28539	1	4992.28539	F(1, 9989)	= 3899.56
Residual	12788.1087	9,989	1.28021911	Prob > F	= 0.0000
Total	17780.3941	9,990	1.77981923	R-squared	= 0.2808
				Adj R-squared	= 0.2807
				Root MSE	= 1.1315

logincom	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
edulevel	.2156922	.003454	62.45	0.000	.2089216 .2224628
_cons	8.799824	.0215411	408.51	0.000	8.757599 8.842048

Figure 1. Model 1 regression analysis

图 1. 模型 1 结果分析

3.2.2. 模型 2 结果分析

模型 2 中加入新的自变量配偶或同居伴侣的最高受教育程度，在图 2 中可以看出模型的 p 值 < 0.05 ，R-squared 的值为 0.3156，较模型一 R 方值变大，这说明 R 方解释力度变大，31.56% 的收入影响来源于居民受教育程度和配偶的受教育水平；除此之外，两个自变量的 p 值都 < 0.05 ，说明自变量对因变量的

影响是显著的，利用 T 检验(图 3)也可以得出加入新的变量之后，模型的效果更好。同样可以得出模型 2 的公式： $y = 8.702983 + 0.1786742 * x_1 + 0.071886 * x_2$ 。意味着这两个自变量和因变量之间呈现正相关关系。

```
. reg logincom edulevel peduclevel
```

Source	SS	df	MS	Number of obs	=	7,863
Model	4229.17459	2	2114.5873	F(2, 7860)	=	1812.05
Residual	9172.29932	7,860	1.1669592	Prob > F	=	0.0000
				R-squared	=	0.3156
				Adj R-squared	=	0.3154
Total	13401.4739	7,862	1.70458839	Root MSE	=	1.0803

logincom	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
edulevel	.1786742	.0058581	30.50	0.000	.1671908 .1901575
peduclevel	.0701886	.005813	12.07	0.000	.0587936 .0815836
_cons	8.702983	.0247071	352.25	0.000	8.65455 8.751415

Figure 2. Model 2 regression analysis
图 2. 模型 2 结果分析

```
. test peduclevel
```

(1) peduclevel = 0

F(1, 7860) = 145.79
Prob > F = 0.0000

Figure 3. T-test
图 3. T 检验

3.2.3. 模型 3 结果分析

模型 3 中加入了新的自变量父亲的受教育程度，从中可以看出整个模型是显著的，R-squared 的值上升，这说明自变量对因变量的解释力度变大；三个自变量的 p 值均 < 0.05，也说明三个自变量对因变量的影响是显著的，并且呈现正相关的关系(图 4)。模型 3 的公式可表示为 $y = 8.66824 + 0.1704935 * x_1 + 0.0608742 * x_2 + 0.0428005 * x_3$ 。

```
. reg logincom edulevel peduclevel feduclevel
```

Source	SS	df	MS	Number of obs	=	7,370
Model	4095.93888	3	1365.31296	F(3, 7366)	=	1207.60
Residual	8328.0166	7,366	1.13060231	Prob > F	=	0.0000
				R-squared	=	0.3297
				Adj R-squared	=	0.3294
Total	12423.9555	7,369	1.68597577	Root MSE	=	1.0633

logincom	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
edulevel	.1704935	.0061151	27.88	0.000	.1585062 .1824807
peduclevel	.0608742	.0059797	10.18	0.000	.0491523 .0725962
feduclevel	.0428005	.0059535	7.19	0.000	.0311299 .0544711
_cons	8.666824	.0255512	339.19	0.000	8.616736 8.716911

Figure 4. Model 3 regression analysis
图 4. 模型 3 结果分析

3.2.4. 模型 4 结果分析

模型 4 中加入了新的变量母亲的最高受教育程度，可以看出整个模型是显著的，各个自变量对因变量的影响也是显著的，p 值均 < 0.05，并且系数都为正，这说明四个字自变量和因变量之间呈正相关关系，R-squared 的值上升，这说明自变量对因变量的解释力度进一步变大(图 5)。模型 4 的公式可表示为 $y = 8.665019 + 0.1672155*x_1 + 0.0564421*x_2 + 0.0284888*x_3 + 0.0388127*x_4$ 。

```
. reg logincom edulevel peduclevel feduclevel meduclevel
```

Source	SS	df	MS	Number of obs	=	7,281
Model	4085.75884	4	1021.43971	F(4, 7276)	=	908.25
Residual	8182.7507	7,276	1.12462214	Prob > F	=	0.0000
				R-squared	=	0.3330
				Adj R-squared	=	0.3327
Total	12268.5095	7,280	1.68523483	Root MSE	=	1.0605

logincom	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
edulevel	.1672155	.0061929	27.00	0.000	.1550756 .1793554
peduclevel	.0564421	.0060483	9.33	0.000	.0445855 .0682986
feduclevel	.0284888	.0069535	4.10	0.000	.0148578 .0421198
meduclevel	.0388127	.0090603	4.28	0.000	.021052 .0565735
_cons	8.665019	.0256107	338.34	0.000	8.614815 8.715224

Figure 5. Model 4 regression analysis

图 5. 模型 4 结果分析

3.2.5. 模型 5 结果分析

模型 5 中加入新的新的变量教育培训支出，图中显示整个模型是显著的，R-squared 的值没有明显上升，说明教育培训支出对居民收入的影响没有很大的解释力度，并且教育培训支出的 p 值 > 0.05，这说明教育培训支出对居民收入的没有影响，因此在后续的分析中就不考虑教育培训支出对居民收入的影响(图 6)。

```
. reg logincom edulevel peduclevel feduclevel meduclevel educmoney
```

Source	SS	df	MS	Number of obs	=	2,245
Model	1265.36904	5	253.073808	F(5, 2239)	=	225.13
Residual	2516.86381	2,239	1.12410175	Prob > F	=	0.0000
				R-squared	=	0.3346
				Adj R-squared	=	0.3331
Total	3782.23285	2,244	1.68548701	Root MSE	=	1.0602

logincom	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
edulevel	.1701533	.0111155	15.31	0.000	.1483556 .1919509
peduclevel	.059298	.0108371	5.47	0.000	.0380462 .0805498
feduclevel	.020126	.0124526	1.62	0.106	-.0042938 .0445458
meduclevel	.0433959	.016583	2.62	0.009	.0108763 .0759155
educmoney	4.54e-08	5.28e-08	0.86	0.390	-5.82e-08 1.49e-07
_cons	8.648393	.0465099	185.95	0.000	8.557186 8.7396

Figure 6. Model 5 regression analysis

图 6. 模型 5 结果分析

为检验各个自变量是否存在共线性问题，因此进行 VIF 检验操作。

由图 7 中可以看出 VIF 的均值为 1.85，并没有明显大于 1，因此自变量的选择不存在共线性问题。

```

. estat vif

```

Variable	VIF	1/VIF
edulevel	2.37	0.422497
peduclevel	2.36	0.423454
meduclevel	1.85	0.539316
feduclevel	1.69	0.592646
educmoney	1.00	0.998297
Mean VIF	1.85	

Figure 7. VIF test

图 7. VIF 检验

从表 3 的结果中，也可以看到，通过嵌套模型的不停增加，除教育培训指出这个变量之外，其他变量都对居民收入都有影响，解释力度在不断变大。

Table 3. Model optimization

表 3. 模型优化

VARIABLES	(1)	(2)	(3)	(4)	(5)
	m1	m2	m3	m4	m5
	logincom	logincom	logincom	logincom	logincom
edulevel	0.216*** (0.003)	0.179*** (0.006)	0.170*** (0.011)	0.167*** (0.006)	0.170*** (0.011)
peduclevel		0.070*** (0.006)	0.059*** (0.011)	0.056*** (0.006)	0.059*** (0.011)
feduclevel			0.020 (0.012)	0.028*** (0.007)	0.020 (0.012)
meduclevel			0.043*** (0.017)	0.039*** (0.009)	0.043*** (0.017)
educmoney			0.000 (0.000)		0.000 (0.000)
Constant	8.800*** (0.022)	8.703*** (0.025)	8.648*** (0.047)	8.665*** (0.026)	8.648*** (0.047)
Observations	9991	7863	2245	7281	2245
R-squared	0.281	0.316	0.335	0.333	0.335

Standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.

4. 结论

通过以上研究分析,可以得出居民年收入会受到受教育程度的影响,并且居民自己的受教育程度、配偶或同居伴侣的受教育程度,父亲母亲的受教育程度与居民的收入呈现正相关关系。我们可以这样进行解释,居民自己的受教育程度高,学识水平高,更容易找到薪资高的工作,这直接会对居民收入产生影响;配偶或同居伴侣的受教育程度高,很大程度会找到高薪工作,这样就为居民自己减轻压力负担,居民自身的收入就会保持较高的水平;同样,父亲母亲的受教育程度高,具备优越家庭的学习氛围,重视对子女的学习教育,直接对子女产生影响,在一定程度上会提高子女的学习素质,从另一方面也说明家境殷实,为后代奠定了财富基础,那么居民个人会遗传父母的财富,居民自身的能力会保持在相对高的水平,也会对居民收入产生正向的影响。从分析来看,我们可以知道居民自身受教育的程度会对居民收入产生直接的正向影响,配偶或同居伴侣、父母的受教育程度对居民收入会产生间接的正向影响。根据分析结果来看,教育培训支出没有发生影响,可能教育支出和收入没有直接的相关关系。我们试图这样进行解释,随着教育培训支出的上升,个人的受教育水平可能会提高,但是另一方面教育支出的扩大可能会导致当下收入的减少,这也说明教育培训支出与收入之间没有直接的联系,因此在模型中教育培训支出不会影响居民收入。

教育是对中华民族伟大复兴具有决定性意义的事业,不管是我们国家还是家庭都应该重视教育的发展,提升受教育水平一方面有助于提升我们自己的认知水平,扩展认识世界的角度,另一方面也有利于家庭水平的提高,有利于国家的发展。因此我们要注重教育水平的提升,进一步强化义务教育,提高全民的文化教育水平,让大部分人集中聚焦于高中阶段的教育水平,扫除文盲;同时随着互联网的普及,相关教育部门可以整合优秀的中高等学校的教育资源,集中优秀师资探索开展“慕课”“翻转课堂”等多种形式的网络教育,让那些没有能力上优秀学校的人也可以得到优质的辅导,从而提升教育水平。在家庭教育中,父母也需要不断的加强学习,提升个人学识和素质,做好家庭教育,重视对孩子学习能力的培养。

参考文献

- [1] 董良. 教育、工作经验与家庭背景对居民收入的影响——对明瑟方程和“布劳—邓肯”模型的综合[J]. 中国社会科学院研究生院学报, 2016(4): 103-109.
- [2] 谢周亮. 家庭背景、人力资本与个人收入差异[J]. 财经科学, 2010(5): 70-76.
- [3] 曹麦. 学校教育、家庭教育与个体收入——基于 CGSS2013 数据的研究[J]. 兰州财经大学学报, 2017, 33(5): 50-55.
- [4] 徐秋艳, 张秋炜. 家庭受教育水平、健康水平与收入水平——基于中国家庭追踪调查的实证分析[J]. 调研世界, 2018(8): 19-28.
- [5] 张津琛, 储宁, 孙东宁. 教育、健康、语言水平对个人收入的影响[J]. 当代经济, 2019(6): 159-161.
- [6] 李秀玉, 蔡玉洁. 质量可比的教育与居民收入关系的实证分析——基于 CGSS 数据的研究[J]. 统计学报, 2020, 1(6): 85-94.