

关于我国普通高等学校招生规模影响因素的分析

摘要

本文以我国普通高等学校招生规模为研究对象，探讨了其影响因素。通过选取人口出生率、居民人均可支配收入、人均 GDP、高校数量、高校专任教师数量以及教育经费这六个变量，建立了计量经济学模型，并且对其进行多重共线性检验、自相关检验和异方差检验。研究结果显示，人口出生率对高校招生规模有显著的负面影响，而高校专任教师数量对招生规模有显著的正面影响。本研究为政府制定教育政策提供了科学依据，有助于高等教育资源的合理配置和教育质量的提升。

关键词：线性回归；多重共线性；自相关；异方差

Abstract

This paper focuses on the enrollment scale of regular higher education institutions in China and explores its influencing factors. By selecting six variables—birth rate, per capita disposable income of residents, per capita GDP, number of higher education institutions, number of full-time teachers in higher education institutions, and education funding—a econometric model was established and subjected to tests for multicollinearity, autocorrelation, and heteroscedasticity. The research results show that the birth rate has a significant negative impact on the enrollment scale of higher education institutions, while the number of full-time teachers has a significant positive impact on the enrollment scale. This study provides a scientific basis for the government to formulate education policies, contributing to the rational allocation of higher education resources and the improvement of education quality.

Keywords: Linear regression; multicollinearity; autocorrelation; heteroskedasticity

1. 研究背景与意义

1.1 研究背景

普通高等教育是国家培养高素质人才、推动科技创新和社会进步的重要基础。在我国，普通高等学校的招生规模不仅关系到教育资源的合理配置和教育质量的提高，更直接影响到社会经济发展和人民生活水平的提升。近年来，随着我国经济的快速发展和人民生活水平的不断提高，普通高等教育的发展也取得了显著成就。然而，招生规模的扩大也伴随着一系列挑战，如教育资源的紧张、教学质量的保障以及毕业生就业压力等问题。此外，近年来，我国的高考制度也一直在改革创新，从 2021 年高考，江苏等地首次实行“3+1+2”政策，到今年贵州、甘肃等 7 省的跟进，高等教育一直收到各界的高度关注。

在人口老龄化趋势明显、社会经济结构调整和产业升级加速的背景下，分析和研究影响我国普通高等学校招生规模的因素具有重要的现实意义。通过定量分析招生规模的影响因素，可以为政府制定相关政策提供科学依据，帮助高等教育机构合理规划招生计划，提高教育资源的利用效率，促进教育公平和质量提升。

1.2 研究意义

- **政策制定依据：**通过对影响普通高等学校招生规模的因素进行系统分析，可以为政府制定教育政策提供有力的支持。例如，了解人口出生率对招生规模的影响，可以帮助政府提前规划未来教育资源的配置。
- **资源合理配置：**研究居民人均可支配收入、人均 GDP 等经济因素对招生规模的影响，可以为高等教育资源的合理配置提供参考，从而提高教育资源的利用效率，减少资源浪费。
- **教育质量提升：**通过分析高校数量、高校专任教师数量、教育经费等对招生规模的影响，有助于了解教育资源投入对教育质量的影响，进而推动教育质量的持续提升。

综上所述，研究我国普通高等学校招生规模的影响因素，不仅具有重要的理论价值，更具有显著的现实意义。通过科学分析，可以更好地推动我国高等教育的健康发展，促进社会经济的持续进步。

2. 因素分析及数据来源

2.1 影响因素分析

人口出生率: 出生率的高低直接决定了未来进入高等教育阶段的学龄人口数量。如果出生率持续下降,将导致适龄学生人数减少,这可能会使得高校面临招生压力,需要调整招生规模和策略,以保持学校的生源数量和质量。

居民人均可支配工资性收入: 居民人均可支配收入的增加通常会提高家庭对教育的投资能力,包括支付更高的学费和其他教育相关费用。这可能导致对高等教育的需求增加,从而促使高等学校扩大招生规模以满足市场需求。

人均 GDP: 随着人均 GDP 的提升,社会对高技能和高学历人才的需求增加,教育需求结构可能会发生变化,这可能导致高等教育的需求增长,促使高等学校扩大招生规模,特别是在科技、工程、管理等领域

高校数量: 高校数量的增加意味着可以提供更多的高等教育席位,这有助于满足社会对高等教育的需求。在高校数量增多的情况下,招生规模可能会相应扩大,以充分利用教育资源。

高校专任教师数量: 专任教师数量的多少决定了高校能够提供的教学资源 and 课程选择的多样性。如果教师资源充足,高校可以提供更广泛的课程选择,吸引更多学生报考。

教育经费: 教育经费的增加可以用于扩建校园设施、增加实验室和图书馆资源等,从而提高学校的教学能力和招生容量。随着教育经费的增加,学校可以开设更多的课程和专业,吸引更多学生选择该校就读,进而可能增加招生规模。

2.2 变量说明

表 1 变量说明

变量	说明	变量	说明
Y	高校招生规模	X_4	高校数量
X_1	人口出生率	X_5	高校专任教师数量
X_2	居民人均可支配工资性收入	X_6	教育经费
X_3	人均 GDP		

2.3 数据及来源

以下数据均来自国家统计局 <https://www.stats.gov.cn/>

表 2 数据集

指标	高校招生数(万人)	人口出生率(‰)	居民人均可支配收入(元)	人均国内生产总值指数	普通高校数(所)	普通高等学校专任教师数(万人)	教育经费(百亿元)
2004	447.3	12.29	3452	12487	1731	85.8	72.43
2005	504.5	12.4	3859	14368	1792	96.6	84.19
2006	546.1	12.09	4426	16738	1867	107.6	98.15
2007	565.9	12.1	5222	20494	1908	116.8	121.48
2008	607.7	12.14	5841	24100	2263	123.7451	145.01
2009	639.5	11.95	6481	26180	2305	129.5	165.03
2010	661.8	11.9	7320	30808	2358	134.3	195.62
2011	681.5	13.27	8313	36277	2409	139.3	238.69
2012	688.8	14.57	9379	39771	2442	144	286.55
2013	699.8	13.03	10411	43497	2491	149.7	303.65
2014	721.4	13.83	11421	46912	2529	153.5	328.06
2015	737.8	11.99	12459	49922	2560	157.3	361.29
2016	748.6	13.57	13455	53783	2596	160.2	388.88
2017	761.5	12.64	14620	59592	2631	163.3	425.62
2018	790.9931	10.86	15829	65534	2663	167.2753	461.43
2019	914.9026	10.41	17186	70078	2688	174.0145	501.78
2020	967.4518	8.52	17917	71828	2738	183.2982	530.34
2021	1001.3151	7.52	19629	81370	2756	186.5541	578.74
2022	1014.5421	6.77	20590	85310	2760	196.3121	613.29
2023	1042.2	6.39	22053	89358	2822	206.0893	/

3 模型建立

3.1 散点图分析

绘制高校招生规模 Y 关于人口出生率 X_1 、居民人均可支配工资性收入 X_2 、人均 GDP X_3 、高校数量 X_4 、高校专任教师数量 X_5 、教育经费 X_6 这六个解释变量的散点图。

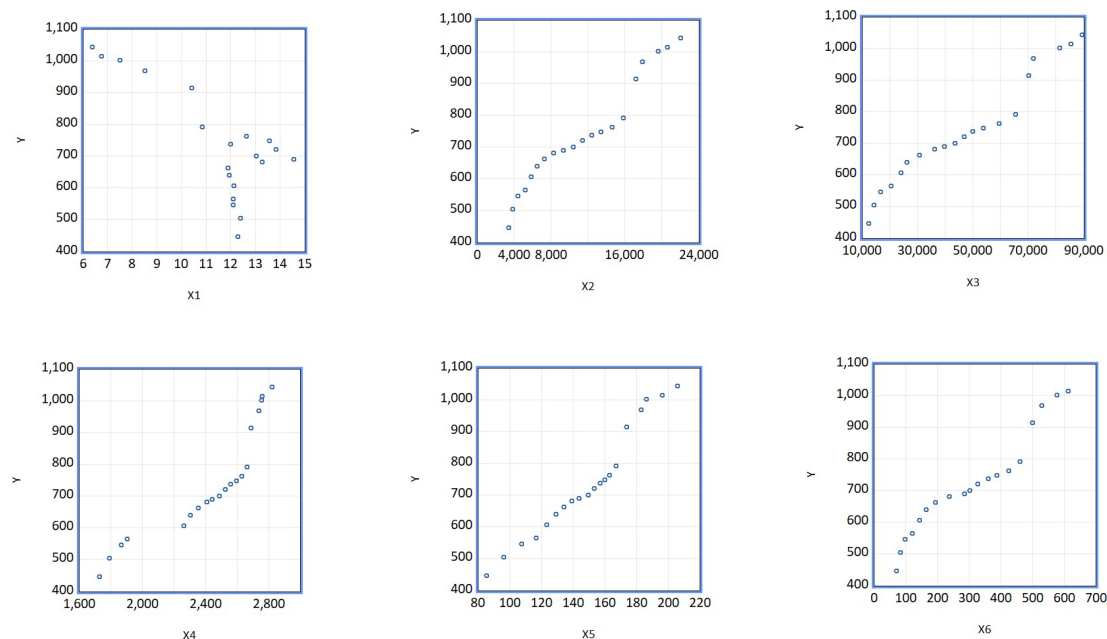


图 1 Y 关于 X 散点图

根据散点图来看，被解释变量 Y 关于各解释变量都存在一定的线性关系

表 3 Eviews 分析结果

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	409.5665	101.4836	4.035789	0.0014
X1	-21.20082	5.384595	-3.937311	0.0017
X2	-0.029045	0.025535	-1.137423	0.2759
X3	-0.001571	0.006848	-0.229435	0.8221
X4	0.017043	0.088196	0.193242	0.8498
X5	3.620309	1.439883	2.514308	0.0259
X6	1.218942	0.734165	1.660311	0.1208
R-squared	0.986971	Mean dependent var		737.1802
Adjusted R-squared	0.980958	S.D. dependent var		173.3657
S.E. of regression	23.92320	Akaike info criterion		9.456792
Sum squared resid	7440.153	Schwarz criterion		9.805298
Log likelihood	-87.56792	Hannan-Quinn criter.		9.524824
F-statistic	164.1325	Durbin-Watson stat		2.395091
Prob(F-statistic)	0.000000			

故此模型建立为线性回归模型。

$$Y = C + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + u$$

根据所给数据，在 Eviews12 中分析之后，得到结果如上述表 3 所示：

$$Y = 409.567 - 21.201X_1 - 0.029X_2 - 0.002X_3 + 0.017X_4 + 3.620X_5 + 1.219X_6$$
$$t = (4.036) \quad (-3.937) \quad (-1.137) \quad (-0.229) \quad (0.193) \quad (2.514) \quad (1.660)$$

$$R^2 = 0.986971 \quad \bar{R}^2 = 0.980958 \quad F = 164.1325 \quad n = 20$$

从结果中可以看出，该模型的 $R^2 = 0.986971$ ，表示该模型解释了约 98.7% 的因变量 Y 的变异，这意味着模型的拟合程度非常高。考虑自变量的个数之后，调整了的 $\bar{R}^2 = 0.980958$ ，仍然表示模型有很高的拟合度。由 F 统计量为 164.1325，P 值为 0.000000，可判断模型整体上显著，“人口出生率”、“居民人均可支配工资性收入”、“人均 GDP”等六个变量联合起来对高校招生规模有显著影响。

但是当考虑单个变量的影响时，当我们取 $\alpha = 0.05$ ， $t_{0.025}(20-6) = 2.145$ ，该结果中只有 X_1 和 X_5 的 t 值大于该临界点（或者 p 值小于 0.05），表明 X_1 和 X_5 对因变量 Y 在 $\alpha = 0.05$ 显著水平上有显著的影响。其余变量的 t 值均小于临界值（p 值大于 0.05），对 Y 的影响在 $\alpha = 0.05$ 显著水平上不显著。此外，通过与散点图进行比较，我们发现 X_2 和 X_3 的系数与预期符号相反。这些说明，该模型可能存在多重共线性等问题，需要进一步的检验。

4. 模型检验及修正

4.1 多重共线性检验

多重共线性会影响回归模型的稳定性和解释力，其存在是因为模型中的两个或多个解释变量高度相关，这会导致回归系数估计值不稳定且不准确。此外，它还会造成回归系数不稳定，预测能力下降、标准误增大，显著性检验失效、回归系数符号和大小不合理等危害，所以我们要对模型进行多重共线性检测

4.1.1 简单相关系数检验法

根据所给的数据，使用 Eviews12 计算皮尔逊相关系数矩阵，所得的结果以及 MATLAB 可视化如下：



图 2 相关系数及可视化

由相关系数矩阵以及可视化结果可以看出，部分解释变量之间的相关系数较高，证实确实存在一定的多重共线性。

4.1.2 方差扩大因子法

根据所给的数据，使用 Eviews12 计算方差扩大因子 VIF，所得的结果如下：

表 4 方差扩大因子

Variable	Coefficient Variance	Uncentered VIF	Centered VIF
C	10298.92	359.9012	NA
X1	28.99386	137.2078	5.254237
X2	0.000652	3773.225	763.2863
X3	4.69E-05	4543.845	935.9734
X4	0.007779	1616.019	30.07866
X5	2.073262	1676.861	73.56136
X6	0.538998	2605.079	603.3541

根据运行的结果来看，解释变量 X_2 、 X_3 、 X_4 、 X_5 、 X_6 的方差扩大因子都大于 10，其中 X_2 、 X_3 、 X_6 的方差扩大因子都远大于 10，这表明存在着严重的多重共线性问题。

4.1.3 多重共线性的修正

有前面的分析检验可知，原模型存在着严重的多重共线性问题，为解决该问题，我们尝试先将各变量按如下进行对数变换：

$$\ln Y = C + \beta_1 \ln X_1 + \beta_2 \ln X_2 + \beta_3 \ln X_3 + \beta_4 \ln X_4 + \beta_5 \ln X_5 + \beta_6 \ln X_6 + u$$

表 5 取对数之后的线性回归结果

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.979344	2.834345	1.756788	0.1025
LOG(X1)	-0.267433	0.070603	-3.787843	0.0023
LOG(X2)	-0.179797	0.318516	-0.564483	0.5820
LOG(X3)	-0.103462	0.466385	-0.221838	0.8279
LOG(X4)	0.038917	0.296392	0.131301	0.8975
LOG(X5)	0.584108	0.273560	2.135211	0.0523
LOG(X6)	0.318774	0.349156	0.912985	0.3779
R-squared	0.987834	Mean dependent var		6.576444
Adjusted R-squared	0.982219	S.D. dependent var		0.236756
S.E. of regression	0.031571	Akaike info criterion		-3.803957
Sum squared resid	0.012957	Schwarz criterion		-3.455451
Log likelihood	45.03957	Hannan-Quinn criter.		-3.735925
F-statistic	175.9218	Durbin-Watson stat		1.780579
Prob(F-statistic)	0.000000			

使用 Eviews 分析得到变换后各解释变量的系数如下图所示

$$\ln Y = 4.979 - 0.267 \ln X_1 - 0.180 \ln X_2 - 0.103 \ln X_3 + 0.039 \ln X_4 + 0.584 \ln X_5 + 0.319 \ln X_6$$

$$t = (1.757) \quad (-3.788) \quad (-0.564) \quad (-0.222) \quad (0.131) \quad (2.135) \quad (0.913)$$

$$R^2 = 0.987834 \quad \bar{R}^2 = 0.982219 \quad F = 175.9218 \quad n = 20$$

从结果中可以看出，经过对数变换处理之后，该模型的 $R^2 = 0.987834$ ， $\bar{R}^2 = 0.982219$ ，仍然表示模型有很高的拟合度。由 F 统计量为 175.9218，P 值为 0.000000，可判断模型整体上显著。然而，虽然 X_3 的系数符号变得与预期一致，但 X_2 的系数仍然与预期符号相反。此外，在 $\alpha = 0.05$ 的显著水平下， $t_{0.025}(20-6) = 2.145$ ，只有 X_1 的 t 值大于临界值，表明对 Y 有显著的影响，其他解释变量均不显著。故对数变换未能很好的解决该模型的多重共线性问题。

下面采用逐步回归方法筛选并剔除引起多重共线性的变量，通过使用 SPSSAU 进行计算，最终保留解释变量人口出生率 X_1 和高校专任教师数量 X_5 。此外，通过检验发现， X_1 和 X_5 的 VIF 值均为 1.765592，小于 10，说明其不再存在多重共线性的情况。

表 6 修正后的结果

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	276.1495	63.06623	4.378722	0.0004
X1	-16.80428	3.129999	-5.368783	0.0001
X5	4.388302	0.223691	19.61771	0.0000
R-squared	0.982868	Mean dependent var		737.1802
Adjusted R-squared	0.980852	S.D. dependent var		173.3657
S.E. of regression	23.98944	Akaike info criterion		9.330585
Sum squared resid	9783.382	Schwarz criterion		9.479945
Log likelihood	-90.30585	Hannan-Quinn criter.		9.359742
F-statistic	487.6462	Durbin-Watson stat		1.597723
Prob(F-statistic)	0.000000			

修正多重共线性后的模型为：

$$Y = 276.150 - 16.804X_1 + 4.388X_5$$

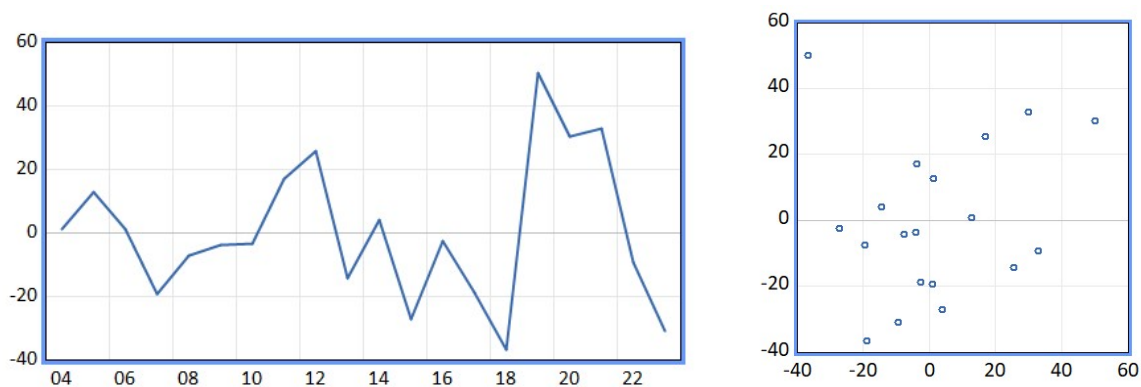
$$t = (4.379) \quad (3.130) \quad (19.618)$$

$$R^2 = 0.982868 \quad \bar{R}^2 = 0.980852 \quad F = 487.6462 \quad n = 20$$

由此可见， X_1 和 X_5 都是比较显著的，表明人口出生率，高校专任教师数量是影响高校招生规模的主要因素。 X_1 和 X_5 系数的符号符合经济意义，表明招生规模随着人口出生率的降低和高校专任教师数量的增加而增大。另外，可决系数和修正的可决系数的值都比较大，表明模型的拟合度比较高。

4.2 自相关检验

4.2.1 图示检验法和 DW 检验法

图 3 残差图与 (e_{t-1}, e_t) 的散点图

修正了多重共线性的问题之后，我们利用新的模型的残差，绘制残差图以及 (e_{t-1}, e_t) 的散点图。从此图中可以看出，残差项 e_t 随着时间 t 的变化虽呈现了锯齿形的形状，但逐次的变化没有呈现明显的规律，可以推测其可能存在相关。

此外，由前面的分析结果可知，该模型的 DW 值为 1.598。根据 DW 检验法，查表可得 $d_L = 1.100$ ， $d_U = 1.537$ 。因为 $d_U < DW < 4 - d_U$ ，可证明该模型的随机误差项不存在自相关性。

4.2.2 Breusch-Godfrey 检验

对于线性回归模型 $Y = C + \beta_1 X_1 + \beta_5 X_5 + u_t$ ，我们假设 u_t 服从各 p 阶（ $p=1、2、3、4$ ）自回归模式：

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \cdots + \rho_p u_{t-p} + v_t$$

通过计算，得到各阶的 TR^2 值如下表所示。从表中可以看出，各 p 阶情况下的 TR^2 值均小于 $\chi^2_\alpha(p)$ 值，进一步证明了该模型不存在自相关性。

表 7 各阶的 TR^2 值

p 阶	TR^2 值	$\chi^2_\alpha(p)$ 值	是否显著
1	0.593107	3.84146	否
2	0.970113	5.99147	否
3	3.381476	7.81473	否
4	7.816032	9.48773	否

4.3 异方差检验

异方差检验是一种常见的检验方法，用于检验数据的方差是否在不同条件下存在显著差异。在建立统计模型时，通常假设数据的方差是恒定的，即方差齐性。然而，在实际应用中，数据可能存在方差不齐性，即方差在不同条件下具有显著的差异。这种方差不齐性可能导致模型的偏误和无效性，因此我们还需要对我们的模型进行异方差检验。

因为模型有 X_1 和 X_5 两项，使用 White 检验时，我们设辅助函数为

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_1^2 + \alpha_3 X_1 X_5 + \alpha_4 X_5 + \alpha_5 X_5^2 + v_t$$

表 8 White 检验结果

Heteroskedasticity Test: White				
F-statistic	2.989658	Prob. F(5,14)		0.0484
Obs*R-squared	10.32758	Prob. Chi-Square(5)		0.0665
Scaled explained SS	5.943692	Prob. Chi-Square(5)		0.3117
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-35876.35	35309.46	-1.016055	0.3268
X1^2	-143.3034	83.54707	-1.715241	0.1083
X1*X5	-9.124041	11.40457	-0.800033	0.4371
X1	4787.024	3737.572	1.280784	0.2211
X5^2	0.170400	0.242934	0.701428	0.4945
X5	82.35854	178.8812	0.460409	0.6523
R-squared	0.516379	Mean dependent var		489.1691
Adjusted R-squared	0.343657	S.D. dependent var		633.4643
S.E. of regression	513.2009	Akaike info criterion		15.56254
Sum squared resid	3687253.	Schwarz criterion		15.86126
Log likelihood	-149.6254	Hannan-Quinn criter.		15.62085
F-statistic	2.989658	Durbin-Watson stat		2.156001
Prob(F-statistic)	0.048419			

从图中的结果可以看出， $nR^2 = 10.32758$ ，由 White 检验知，在 $\alpha = 0.05$ 下，查找 χ^2 分布表，得到临界值 $\chi^2_{0.05}(5) = 11.0705$ ，因为 $nR^2 < \chi^2_{0.05}(5)$ ，所以接受原假设，表明模型中随机误差不存在异方差。

5 结论与建议

经过前面的检验和修正之后，最终得到的模型为：

$$Y = 276.150 - 16.804X_1 + 4.388X_5$$
$$t = (4.379) \quad (3.130) \quad (19.618)$$

$$R^2 = 0.982868 \quad \bar{R}^2 = 0.980852 \quad F = 487.6462 \quad n = 20$$

该模型的 $R^2 = 0.982868$ ， $\bar{R}^2 = 0.980852$ ，仍然表示模型有很高的拟合度。且各解释变量参数的 t 检验均显著，不存在多重共线性、自相关性和异方差。通过分析该模型，我们可以得出一些结论，并且根据结论提出一些建议。

5.1 结论

人口出生率的对高校的招生规模具有负面影响:模型中 x_1 的系数 -16.804 表明人口出生率对普通高等学校招生规模具有显著的负面影响。这意味着在其他条件不变的情况下,人口出生率每增加一个单位,普通高等学校的招生规模将减少约 16.804 个单位。这可能是由于人口出生率较高时,年轻人口多,基础教育压力大,导致资源更多地分配到基础教育层面,而高等教育资源相对紧张。

高校专任教师数量对高校招生规模具有正面影响:模型中 x_5 的系数 4.388 表明高校专任教师数量对普通高等学校招生规模具有显著的正面影响。即在其他条件不变的情况下,每增加一个单位的高校专任教师数量,普通高等学校的招生规模将增加约 4.388 个单位。这说明师资力量增强能够有效扩大招生规模,可能是由于教师数量的增加改善了教学质量和教育资源的供给能力,从而吸引更多学生入学。

5.2 关于高校招生规模的对策建议

优化生育政策:由于人口出生率对招生规模有负面影响,政府应综合考虑生育政策的影响,合理规划各阶段教育资源的分配。可以通过加强基础教育与高等教育的衔接,确保在高人口出生率地区也能有足够的高等教育资源支持。

加大对高校教师的投入:模型表明,高校专任教师数量对招生规模有显著的正面影响。建议政府和高校加大对教师队伍建设的投入,通过提高教师待遇、加强师资培训和引进优秀人才等方式,提升高校的教学水平和吸引力,从而扩大招生规模。

加强教育资源的统筹规划:各地政府应根据当地人口出生率的变化趋势和经济发展水平,统筹规划教育资源的配置。特别是在人口出生率较高的地区,应提前规划教育资源的投入,确保高等教育能够适应未来人口变化的需求。