Advances in Social Sciences 社会科学前沿, 2023, 12(1), 49-58
Published Online January 2023 in Hans. https://www.hanspub.org/journal/ass
https://doi.org/10.12677/ass.2023.121007

Hans汉斯

皮尔逊相关系数在高等教育发展中的影响因素的研究

——以长三角地区为例

方　澍，唐　斌，傅周超，骆晨丹，胡金杰

绍兴文理学院，浙江　绍兴

收稿日期：2022年11月22日；录用日期：2023年1月9日；发布日期：2023年1月16日

摘　要

长三角地区的高等教育虽然在中国处于相对领先地位，但是其高等教育的发展水平呈现出不平衡的现状，我们有必要对高等教育的影响因素进行研究。通过量化分析长三角高等教育条件，在经过夏皮洛威尔克法通过正态性检验之后，计算皮尔逊相关系数研究影响其发展的主要因素。最后得到的结论是：在校学生资源、教师资源以及教育经费，这几项指标对高等教育发展的影响是最显著的。

关键词

皮尔逊相关系数，正态分布检验，高等教育影响因素

A Study of the Factors Influencing Pearson's Correlation Coefficient in the Development of Higher Education

—Taking the Yangtze River Delta as an Example

Shu Fang, Bin Tang, Zhouchao Fu, Chendan Luo, Jinjie Hu

Shaoxing University, Shaoxing Zhejiang

Received: Nov. 22nd, 2022; accepted: Jan. 9th, 2023; published: Jan. 16th, 2023

Abstract

Although the higher education in the Yangtze River Delta region is in a relatively leading position

文章引用: 方澍, 唐斌, 傅周超, 骆晨丹, 胡金杰. 皮尔逊相关系数在高等教育发展中的影响因素的研究[J]. 社会科学前沿, 2023, 12(1): 49-58. DOI: 10.12677/ass.2023.121007

in China, the development level of its higher education is unbalanced. It is necessary for us to study the influencing factors of higher education. Through quantitative analysis of higher education conditions in the Yangtze River Delta, after passing the normality test by Charpiro Wilke's method, calculating the Pearson correlation coefficient studies the main factors influencing its development. The final conclusion is: Resources for current students, teacher resources and funding for education, these indicators have the most significant impact on the development of higher education.

Keywords

Pearson Correlation Coefficient, Normal Distribution Test, Influencing Factors of Higher Education

1. 引言

《中国教育现代化 2035》明确提出，一流的人才培养与创新能力是衡量教育现代化水平的重要标准。现如今，中国高等教育已经迈向普及化阶段，但是在与世界先进水平的较量之中，中国的高等教育一体化发展水平还不高[1]。在短短十年间，长三角地区一体化发展的进程不断加快，在促进人的全面发展、缩小区域发展差距方面取得了举世瞩目的成就，但在极高人类发展水平组中，仍处于相对较低的水平；教育水平或将成为长三角地区人类发展的主要短板[2]。长三角地区作为中国高等教育一体化程度较好的区域，同样面临着高等教育发展水平不平衡的状态，那么就会引发我们的思考：究竟哪些是影响高等教育发展的主要因素？因此，本文通过正态分布检验和皮尔逊相关系数等方式来对高等教育发展的主要影响因素进行研究，并对结论进行一定的分析。

2. 研究准备

2.1. 研究意图

本研究旨在了解长三角地区高等教育发展的不平衡现状，通过收集一系列的数据，进行整理和综合分析，深入了解各个因素与高等教育发展之间的关系，从中筛选出影响最大的几个因素。在得到研究结果之后，为长三角高等教育的发展提供一些建议，提升其高等教育区域一体化的进程。

2.2. 研究思路

首先，将收集的相关数据绘制成三线表，见表 1。根据表格内容，表格中的指标名称即为高等教育发展的影响因素，一共有 21 个影响因素。将教育发展指数看作高等教育发展的评价指标，通过皮尔逊相关系数，刻画各项指标之间的线性关系，通过比较分析它们之间的关系，进而研究得到高等教育发展的主要影响因素。

2.3. 数据来源

表 1 的各项数据是经过整理所得，原始数据来源于：中华人民共和国教育部官方网站(http://www.moe.gov.cn/jyb_sjzl/moe_560/jytjsj_2019/gd/)、中华人民共和国统计局、中国统计年鉴(2014~2020)、中国教育经费统计年鉴(2014~2019)、中国科技统计年鉴(2014~2019)。

Table 1. From 2013 to 2018/2019, the average value of higher education indicators in the Yangtze River Delta region was compared with the national ratio
表 1. 2013~2018/2019 年长三角地区高等教育各项指标均值与全国比值

指标名称	上海	江苏	浙江	安徽
人口	1.75%	5.79%	4.07%	4.49%
GDP	3.71%	10.20%	6.27%	3.35%
本科学校比重	1.22	0.99	1.16	0.8
每十万人口高等教育平均在校学生数	1.33	1.16	0.93	0.88
在校研究生比重	3.19	1.25	0.92	0.63
在校博士生比重	3.86	1.24	0.92	0.49
副高及以上专任教师比重	1.19	1.12	1.07	0.85
正高级专任教师比重	1.46	1.07	1.14	0.69
硕士及以上学位专任教师比重	1.37	1.09	1.1	1
博士学位专任教师比重	2.19	1.36	1.33	0.7
图书数量	2.92%	6.98%	4.28%	3.63%
教学、科研仪器设备资产价值	5.91%	8.35%	4.92%	3.15%
教室数量	2.32%	6.72%	3.76%	4.30%
教育经费支出	5.93%	7.51%	4.93%	2.92%
生均教育事业费支出	1.7	0.97	0.97	0.7
生均公用经费支出	2.25	0.98	0.95	0.83
R & D 人员数量	5.27%	7.28%	5.54%	3.21%
R & D 经费内部支出数量	8.51%	9.11%	5.23%	2.77%
发表科技论文数量	6.40%	9.34%	3.81%	3.08%
出版科技著作数量	6.32%	7.08%	4.24%	2.67%
专利申请数量	5.03%	15.49%	7.53%	4.15%
起点发展指数	27.222	22.144	17.223	17.205
过程发展指数	25.127	18.525	16.104	12.765
结果发展指数	24.475	20.953	18.545	16.268

2.4. 数据解释

其中，高等学校(机构)在校生数 = 普通本专科在校生数 + 高等学校(机构)研究生在校生数。表 1 中人口、GDP、图书数量等含有百分号的数据，它是将长三角各地区的指标均值除以全国的指标总和，因而这些数据是含有百分号并且数值一定小于 100%。而表 1 中本科学校比重、每十万人口高等教育平均在校学生数、在校研究生比重等数据并不含有百分号，这是因为这些指标本身是比重、生均等含义，对于长三角地区来说它们是一个含百分号并小于 100%的数，而当这些数据除以全国时，我们可以理解为两个含百分号的数相除，得到的是一个不含百分号的数；当这个比值大于 1 时，则说明在这项指标中，该地区优于全国的平均水平，当比值小于 1 时，则说明该指标未达到全国的平均水平。

我们选择了近十年的数据绘制成表 1，其中近三年由于某种特殊原因，导致部分指标的数据波动较大，对我们的研究不具太大的参考性，故选择剔除。最终选择 2013 年至 2019 年的各项指标均值与全国的比值作为我们研究高等教育发展主要影响因素的方向与来源。

2.5. 高等教育各项指标的社会重要性

表中所有的指标我们可以依次分类为：宏观指标、高校资源、在校生、教师资源、高校资产、教育经费以及科学研究与实验发展。

其中，宏观指标包括人口和 GDP，这两个指标从宏观角度反映了一个地区的经济发展的总体水平。高校资源包括本科学校比重和每十万人口高等教育平均在校学生数，这两个指标反映了一个地区高等教育的平均水平以及受教育的程度。在校生包括在校研究生比重和在校博士生比重，这两个指标反映了地区高校人才的发展情况，也是高校教学水平与能力的体现。教师资源包括副高及以上专任教师比重、正高级专任教师比重、硕士及以上学位专任教师比重、博士学位专任教师比重这四项，教师资源往往是一个地区教育好差的重要表现因素，优秀且庞大的教师队伍将促进当地教育的持续快速发展。教育经费包括教育经费支出、生均教育事业费支出以及生均公用经费支出这三项，教育经费的多少不仅展现了当地政府对教育的重视程度，也是学校地位与综合实力的体现。科学研究与实验发展包括 R & D 人员数量、R & D 经费内部支出数量、发表科技论文数量、出版科技著作数量以及专利申请数量这五项，科学研究与实验发展是高校教育在科研能力方面的体现，包括学校自身科研的综合能力以及培养学生科学研究的能力与情况。

3. 研究过程与步骤

3.1. 正态分布检验

正态分布检验，是判断一样本所代表的背景总体与理论正态分布是否没有显著差异的检验，具有最重要的意义，也是应用最为广泛的检验方法，是参数统计分析的前提[3]。

皮尔逊相关系数要求样本数据满足正态分布的要求，因此，我们首先需要对样本数据的正态分布进行检验。根据国家标准《数据的统计处理和解释——正态性检验》[4]中规定：样本数量 $n\in[8,50]$ 宜采用夏皮洛－威尔克(Shapiro-Wilk)检验方法[5]。由于本研究中的样本数量较少，因此选用夏皮洛－威尔克检验。

夏皮洛－威尔克检验是基于次序统计量对它们期望值的回归，检验统计量为样本次序统计量线性组合的平方与通常的方差估计量的比值，它是一种完全样本的方差分析形式的检验。具体检验步骤为：

1) 提出两个互相对立的假设。原假设 H_0：随机变量符合正态分布；备择假设 H_1：随机变量不符合正态分布。

2) 将数据点到拟合直线的距离 d 按照升序排列为 $X_{(1)}\le X_{(2)}\le\cdots\le X_{(n)}$。

3) 设 $\left(Z_{(1)},\cdots,Z_{(n)}\right)$ 是来自 $N(0,1)$ 中的样本，$\left(Z_{(1)},\cdots,Z_{(n)}\right)$ 是其顺序统计量。记 $c_{(k)}=E\left(Z_{(k)}\right)$ $(k=1,\cdots,n)$，$c=\left(c_1,\cdots,c_n\right)'$，则有：

当 n 为偶数时，有

$$-c_k=c_{n+1-k},\ k=1,\cdots,\frac{n}{2}$$

当 n 为奇数时，有

$$-c_k=c_{n+1-k},\ k=1,\cdots,\frac{n}{2},\ c_{\frac{(n+1)}{2}}=0$$

记

$$a_k(W)=-\frac{c_k}{\|c\|},\ k=1,\cdots,\left[\frac{n}{2}\right]$$

其中 $\|c\|=\left(c'c\right)^{\frac{1}{2}}$。

针对样本 X 记统计量 W 为

$$W=\frac{\left\{\sum_{k=1}^{\left[\frac{n}{2}\right]}a_k(W)\left[X_{(n+1-k)}-X_{(k)}\right]\right\}^2}{\sum_{k=1}^{n}\left(X_{(k)}-\bar{X}\right)^2}$$

上述等式中的 $a_k(W)$ 可以在国家标准《数据的统计处理和解释——正态性检验》附表中查询到，对于给定的显著性水平 α，当 $W>W_\alpha$ 时，接受原假设 H_0，否则就拒接原假设 H_0。

该过程我们可以通过利用 SPSS 软件进行操作与数据的运行，最终将所得数据汇总成三线表，见表 2。

Table 2. Test for normality
表 2. 正态性检验

指标名称	统计	自由度	显著性
人口	0.952	4	0.730
GDP	0.881	4	0.342
本科学校比重	0.940	4	0.653
每十万人口高等教育平均在校学生数	0.920	4	0.536
在校研究生比重	0.820	4	0.144
在校博士生比重	0.810	4	0.122
副高及以上专任教师比重	0.905	4	0.458
正高级专任教师比重	0.976	4	0.876
硕士及以上学位专任教师比重	0.855	4	0.244
博士学位专任教师比重	0.945	4	0.685
图书数量	0.890	4	0.381

Continued

教学、科研仪器设备资产价值	0.990	4	0.956
教室数量	0.963	4	0.799
教育经费支出	0.995	4	0.983
生均教育事业费支出	0.854	4	0.240
生均公用经费支出	0.718	4	0.019
R & D 人员数量	0.968	4	0.829
R & D 经费内部支出数量	0.913	4	0.496
发表科技论文数量	0.927	4	0.576
出版科技著作数量	0.946	4	0.692
专利申请数量	0.843	4	0.204
起点发展指数	0.864	4	0.274
过程发展指数	0.966	4	0.815
结果发展指数	0.987	4	0.943

在对样本数据进行正态分布检验后，我们发现：夏皮洛－威尔克检验的显著性除“生均公用经费支出”这一指标外均大于 0.05 水平，所以接受原假设，认为该样本符合正态分布的要求。

3.2. 皮尔逊相关系数

皮尔逊相关系数[6]是将两个数据集合标准化剔除量纲影响后的协方差，在衡量两个正态连续变量的线性关系的相关程度上有广泛的运用，计算公式为

具体计算步骤如下：

$$r_{XY} = \frac{Cov(X,Y)}{S_X S_Y} \tag{1}$$

其中，$Cov(X,Y)$ 为样本 X，Y 的协方差。

$$Cov(X,Y) = \frac{\sum_{i=1}^{n}(X_i - EX)(Y_i - EY)}{n-1} \tag{2}$$

EX，EY 为样本 X，Y 的均值。

S_X，S_Y 分别为 X，Y 的样本标准差。

$$S_X = \sqrt{\frac{\sum_{i=1}^{n}(X_i - EX)^2}{n-1}} \tag{3}$$

$$S_Y = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - EY)^2}{n-1}} \tag{4}$$

在上一步骤中，我们已经通过了样本的正态性检验。这一步骤中，我们通过 SPSS 软件继续将数据进行运算与分析，计算得到皮尔逊相关系数并将其绘制成三线表，见表 3。

Table 3. Table of correlation coefficients
表 3. 相关系数表

指标名称	起点发展指数	过程发展指数	结果发展指数
人口	−0.5821	−0.6741	−0.5807
GDP	−0.0096	−0.0140	0.0986
本科学校比重	0.5614	0.7716	0.7438
每十万人口高等教育平均在校学生数	0.9895	0.9624	0.9809
在校研究生比重	0.9471	0.9686	0.9374
在校博士生比重	0.9369	0.9651	0.9308
副高及以上专任教师比重	0.7763	0.8857	0.9111
正高级专任教师比重	0.7952	0.9329	0.9182
硕士及以上学位专任教师比重	0.8965	0.9663	0.9284
博士学位专任教师比重	0.8880	0.9816	0.9653
图书数量	−0.1128	−0.1783	−0.0604
教学、科研仪器设备资产价值	0.5438	0.5224	0.6191
教室数量	−0.2859	−0.4108	−0.2987
教育经费支出	0.6119	0.6237	0.7086
生均教育事业费支出	0.9072	0.9737	0.9394
生均公用经费支出	0.9010	0.9305	0.8842
R & D 人员数量	0.3804	0.4320	0.5237
R & D 经费内部支出数量	0.8093	0.8107	0.8740
发表科技论文数量	0.6281	0.5441	0.6403
出版科技著作数量	0.7758	0.7666	0.8373
专利申请数量	0.0875	0.0468	0.1635

3.3. 计算 p 值

原假设 H_0： $r=0$；备择假设 H_1： $r\neq 0$。

由于本文只考虑教育发展指数受其他指标的影响程度，故只需要排出相关系数表中每行的数据即可。计算 p 值也就是计算各个指标的显著性，其中需要计算的 p 值包括各项指标与起点发展指数、过程发展指数以及结果发展指数相关系数的显著性。表格中的数据我们通过 MATLAB 软件中的函数 corrcoef 求出，最终将所有数据整理并绘制成三线表，见表 4。

Table 4. p-value table
表 4. p 值表

指标名称	起点发展指数	过程发展指数	结果发展指数
人口	0.4179	0.3259	0.4193
GDP	0.9904	0.9860	0.9014
本科学校比重	0.4386	0.2284	0.2562
每十万人口高等教育平均在校学生数	**0.0105	**0.0376	**0.0191
在校研究生比重	*0.0529	**0.0314	**0.0626
在校博士生比重	*0.0631	**0.0349	*0.0692
副高及以上专任教师比重	0.2237	0.1143	*0.0889
正高级专任教师比重	0.2048	*0.0671	*0.0818
硕士及以上学位专任教师比重	0.1035	**0.0337	*0.0716
博士学位专任教师比重	0.1120	**0.0184	**0.0347
图书数量	0.8872	0.8217	0.9396
教学、科研仪器设备资产价值	0.4562	0.4776	0.3809
教室数量	0.7141	0.5892	0.7013
教育经费支出	0.3881	0.3763	0.2914
生均教育事业费支出	*0.0928	**0.0263	*0.0606
生均公用经费支出	*0.0990	*0.0695	0.1158
R & D 人员数量	0.6196	0.5680	0.4763
R & D 经费内部支出数量	0.1907	0.1893	0.1260
发表科技论文数量	0.3719	0.4559	0.3597
出版科技著作数量	0.2242	0.2334	0.1627
专利申请数量	0.9125	0.9532	0.8365

4. 结果分析

4.1. 表格中带*数据的解释

首先我们对于表 4 中所出现的带*的数据进行一个简单的解释。***p < 0.01，在 99%的置信水平上拒绝原假设，**p < 0.05，在 95%的置信水平上拒绝原假设，*p < 0.1，在 90%的置信水平上拒绝原假设。

4.2. 研究结果与分析

当 p 值过大时(p > 0.1)，我们认为结果发生的可能性太小，不可能有任何能够显示出这个结果的大规

模实验。其主要意思为：p 值不显著，即使真实的情况是原假设是错误的，也会因为这种情况发生的概率太小，使得在实际中没办法做那么大样本的实验来支持这一论点。因此在我们的研究中，p 值过大意味着该项指标对长三角地区高等教育的发展影响不大，即该项指标不是长三角地区高等教育发展的主要影响因素。所以我们只需要找出 p 值比较小的数据。

结合表 4 中的所有数据，我们将 $p < 0.01$ 的数据前标上***，$p < 0.05$ 的数据前面标上**，$p < 0.1$ 的数据前标上*，$p > 0.1$ 的数据前不进行标记，以此将所有数据有层次地区分出来。而带有*号的数据所对应的指标即为我们的研究结果，这些指标意味着它们对长三角地区高等教育的发展影响较大，即这些指标是长三角地区高等教育发展的主要影响因素。

研究结果表明，该模型在 90%的置信区间上，每十万人口高等教育平均在校学生数、在校研究生比重、在校博士生比重、副高及以上专任教师比重、正高级专任教师比重、硕士及以上学位专任教师比重、博士学位专任教师比重、生均教育事业费支出以及生均公用经费支出这几项指标是显著的，可以拒绝原假设，因此这几项指标对教育发展具有较大影响，是长三角地区高等教育发展的主要影响因素。

其中，每十万人口高等教育平均在校学生数、在校研究生比重、在校博士生比重总称为在校学生资源；副高及以上专任教师比重、正高级专任教师比重、硕士及以上学位专任教师比重、博士学位专任教师比重这四项指标总称为教师资源；生均教育事业费支出和生均公用经费合称为教育经费。

因此，我们对研究结果进行分析得到：在校学生资源、教师资源以及教育经费，这几项指标是显著的，其对长三角地区高等教育的发展影响最大。

5. 关于长三角地区高等教育发展的建议

5.1. 总体简述

基于 2013 年至 2019 年间的各项数据，从区域整体来看，长三角地区高等教育在我国重要高等教育集群中依然保持领先态势[7]。但是在极高人类发展水平组中，仍处于相对较低的水平。通过上述的研究与分析，我们得到了影响长三角地区高等教育发展的主要因素，因此我们将从在校生资源、教师资源以及教育经费这三方面，为长三角地区高等教育发展提出相应的合理的一些建议。

5.2. 在校学生资源方面

对于高校来说，在校学生是潜在的科研人才，是促进高等教育发展的关键因素。本研究得到：每十万人口高等教育平均在校学生数、在校研究生比重、在校博士生比重，是影响高等教育发展的重要因素。为此建议推动长三角地区高等教育在招生考试制度、人才培养模式、办学体制等方面的改革实现重大突破。需要提高每十万人口高等教育平均在校学生的数量，提高在校硕士生和博士生的比重。要突出在校学生对于高等教育发展的关键作用，培养在校学生的基础知识与专业素养，为学生提供在实践中成长的机会，使其成为高等教育事业发展的后备军。

5.3. 教师资源方面

“师者，所以传道受业解惑也”。教师对学生的影响是十分深刻且长久的，同时也会在很大程度上影响地区高等教育的发展。本研究得到，副高及以上专任教师比重、正高级专任教师比重、硕士及以上学位专任教师比重、博士学位专任教师比重，是影响高等教育发展的重要因素。因此，建议组建高校雄厚的师资队伍，更多地聘请副高、正高级，硕士、博士学位的专任教师。面向社会广泛吸纳优秀的、具有科研能力、能够承担产学研相结合的教师，提高人力资源质量。在此基础上，调动广大科研人员的积极性、主动性和创造性[8]。为推动长三角地区高等教育的发展注入新的生命力量。

5.4. 教育经费方面

高等教育的发展势必离不开经济、政治上的支持，教育经费对高等教育发展的影响是最为重大的，没有足够的教育经费，一个地区的高等教育举步维艰。因此，政府或是地区需要加大教育投资力度，以提高生均教育事业费和生均公用经费，以此来保证高等教育事业的稳步快速发展，这是促进高等教育发展的有效途径。

6. 结语

对数据进行整理与分析，选择夏皮洛－威尔克法通过正态性检验，再计算皮尔逊相关系数的方法，我们研究得到了长三角地区高等教育发展的影响因素。最终我们发现：在校学生资源、教师资源以及教育经费这几个因素对于长三角地区高等教育发展的影响最大。其中，在校学生资源包括每十万人口高等教育平均在校学生数、在校研究生比重和在校博士生比重这三部分；教师资源包括副高及以上专任教师比重、正高级专任教师比重、硕士及以上学位专任教师比重和博士学位专任教师比重这四部分。教育经费包括生均教育事业费支出和生均公用经费这两部分。据此，想要促进长三角地区高等教育的发展，要推动招生考试制度、人才培养模式等方面的改革，提升在校学生资源；要组建高校雄厚的师资队伍，提高人力资源质量；要加大教育投资力度，增加教育经费。

致　谢

最后，感谢学校老师提供的悉心指导和积极帮助，同时也对参考文献中的思想和方法的所有者表示最真挚的谢意。

参考文献

[1] 张继平, 邓可. 长三角高等教育一体化高质量发展的现实困境与路径选择——基于区域创新体系的视角[J]. 长江大学学报(社会科学版), 2022, 45(4): 117-124.

[2] 中国教育在线. 长三角迈入“极高人类发展水平”,教育竞成短板[EB/OL]. https://baijiahao.baidu.com/s?id=1718365318418189983&wfr=spider&for=pc, 2021-12-06.

[3] 刘应成. 考试系统中成绩正态分布检验的设计与实现[J]. 重庆工学院学报, 2004, 18(6): 188-191.

[4] GB4882-2001, 数据的统计处理和解释正态性检验[S]. 北京: 中国标准出版社, 2001.

[5] 梅卫锋. PHC 管桩的单桩竖向承载力下限解研究[J]. 路基工程, 2020(2): 93-98.

[6] 李姝昊, 王晓丹, 宋亚飞. 基于皮尔逊系数和不确定测度的冲突证据组合方法[J]. 电子测量与仪器学报, 2021, 35(8): 38-45.

[7] 王新凤, 罗启轩, 钟秉林. 长三角地区高等教育协同发展的历史进程与发展态势[J]. 江苏高教, 2021(9): 1-10.

[8] 刘美凤. 长三角地区高等教育与经济协调发展研究[D]: [硕士学位论文]. 南京: 南京财经大学, 2011.