Public Education Levels and Expenditure: A Correlation Study

Juan José Garcia
jjgarcia@cs.unc.edu

Smiti Kaul
smiti@cs.unc.edu

Sandeep Kumar
sandeep@cs.unc.edu

Saumya Ray
shree@cs.unc.edu

ABSTRACT

Countries prioritize national expenditure in various sectors of citizen issues based upon evaluations that judge the impact and value of such public funding. Education is one such sector that shapes citizens' short- and long-term self-development and societal contribution and upon which public spending exerts significant influence by determining educational access and resources for a national government's constituents. Along these lines, this work evaluates the correlation between public expenditure on education and educational enrollment levels across low, middle, and high income countries. We implement clustering via a latent variable model upon data provided by the United Nations with the hypothesis that countries with greater public funding for education experience higher enrollment rates. Our results support this hypothesis.

KEYWORDS

public expenditure, education, enrollment, clustering, machine learning

1 INTRODUCTION

National wealth, specifically public expenditure on education, and its consequences in student enrollment have been investigated in numerous studies across a wide range of countries [4, 6, 10]. These countries across the globe can be categorized at three financial levels: low, middle, and high income based upon whether their gross domestic product (GDP) measures are low, middle, and high respectively. Countries with greater wealth as exhibited in their GDP measures possess a greater capacity to support a variety of national issues including education.

Low and middle income countries have a smaller spending power than high income countries, so it follows that their public spending budget available for education initiatives is smaller as well. This reality has serious implications, for national- and person-level financial circumstances have been found to influence educational access and resources across the globe. Bergh and Fink noted that only one percent of secondary school students in Ethiopia and Kenya advance to higher education while the over seventy percent advance in higher income countries such as South Korea, Finland, the United States [3]. Developing nations such as some in Sub-Saharan Africa where educational expenses are lower possess 25% of the world's uneducated population [7]. In Peru, poor children are less likely to enroll in elementary schools, and educational participation is further inhibited by supplementary costs associated with transportation and school supplies [9].

Existing research supports the investment in public funding for education, and public investment in educational resources and schools also has been found to increase not only enrollment but academic achievement as well [2, 5]. National level improvements in both academic enrollment and achievement has important positive consequences that motivate a correlation study such as ours. Enrolled students gain employable skills, societal values, and the ability to contribute to their national economic and social welfare [7]. Furthermore, the world's enrollment student population is only increasing and it demands greater attention by the government. In the African continent for example, the number of high education students tripped between 1991 and 2006 [1], and a study in the United States found that enrollment in government-funded community colleges increased by roughly half a million students between 2000 and 2005 and again between 2005 and 2009 [8].

Repeatedly, it has been posited that greater economic wealth measured by GDP and a greater public investment in education benefits a country's state of education [10]. We focus on the percentage of a country's national population

COMP 562, The University of North Carolina at Chapel Hill, USA

that is enrollment in school and aim to analyze a correlation between this measure and a country's GDP and expenditure on education. This paper is organized as follows: Section 1 provides the background and motivation for our analysis; Section 2 explains our data; Section 3 details our clustering methodology based upon latent variable modeling; and Section 4 and 5 present our results and conclusion, respectively.

2 DATA

We used two datasets provided by the United Nations at https://data.un.org/. The "Public expenditure on education" dataset provides the following information on all countries between 2002 and 2017: expenditure type, level of education, total government expenditure, and GDP. The "Enrollment in primary, secondary, and tertiary education levels" dataset provides the following information on all countries for the same time period: number of student enrolled in primary, secondary, and tertiary education levels and gross enrollment values. We combined this information to form our single, master dataset described in the Table 1.

3 METHODOLOGY

Our analysis aims to evaluate the relationships between a country's gross domestic product (GDP) per capita, public expenditure on education, and educational enrollment levels across low, middle, and high income countries. We implement clustering to understand these relationships by forming our clusters based upon the percentage of a country's population that is enrolled in primary (x_1), secondary (x_2), and tertiary (x_3) education. To this end, we use a latent variable model. More specifically, we use a Beta mixture model, where the likelihood is composed of Beta densities and the prior is categorical. Our (1) likelihood, (2) marginal, (3) prior, and (4) posterior models are specified below. We minimize the negative log likelihood to train our model. We chose to model our data using the Beta distribution because our features such as the percentage of GDP used for public funding on education are contained in a range between 0 and 1. Our latent variables are multiples vectors of parameters such as GDP and percentage expended on education that encapsulate all information about education public spending in low, middle, and high income country groups. Each vector parametrizes a density function that scores the likelihood that a country belongs to the low, middle, or high income group. Furthermore, each cluster has its own Beta density. Implicit in our model are a few assumptions. We assume that all countries have the same age, gender, and other social ratios that may confound our data on enrollment rates.

Likelihood

$$p(x_1^{(i)}, x_2^{(i)}, x_3^{(i)}|k) = \prod_{j=1}^{3} Beta(a_{jk}, b_{jk}) \tag{1}$$

Marginal

$$p(x_1^{(i)}, x_2^{(i)}, x_3^{(i)}) = \sum_{k=1}^{k} \prod_{j=1}^{3} Beta(a_{jk}, b_{jk})P(k) \tag{2}$$

Prior

$$P(k) = Cat(k) \tag{3}$$

Posterior

$$p(k|x_1^{(i)}, x_2^{(i)}, x_3^{(i)}) = \frac{p(x_1^{(i)}, x_2^{(i)}, x_3^{(i)}|k)P(k)}{\sum_c p(x_1^{(i)}, x_2^{(i)}, x_3^{(i)}|c)P(c)} \tag{4}$$

We set up our model to yield three clusters distinguished by country education enrollment levels. We verify the cluster membership and semantic meaning using the p-values described in the Results section. To compute the p-value for each cluster, we define three intervals (low, middle, and high GDP) for each cluster. In a given cluster, the interval with the lowest p-value defines that cluster membership as being formed mainly of countries of low, middle, or high GDP. These intervals are listed in tables 2 and 3. In the next section, we visualize and discuss how the clusters are distributed across GDP per capita, public expenditure, and enrollment levels as discus.

Table 2: P-values for GDP Per Capita Intervals

	< 3,500	3,500 - 45,000	> 45,000
Cluster 1	3.47e-12	9.9e-1	9.92e-1
Cluster 2	9.96e-1	2.07e-2	4.37e-2
Cluster 3	9.99e-1	2.41e-3	5.59e-2

Table 3: P-values for % Expenditure Intervals

	< 3.2%	3.2 - 5.4%	> 5.4%
Cluster 1	2.39e-3	1.25e-1	9.93e-1
Cluster 2	9.60e-1	8.66e-1	9.50e-4
Cluster 3	5.25e-1	2.41e-1	4.35e-1

Table 1: Data Features

Column / Feature Name	Description
Country	Country name
Year	A value between 2002 and 2017
Public expenditure on primary education	% of total government expenditure on education
Public expenditure on secondary education	% of total government expenditure on education
Public expenditure on tertiary education	% of total government expenditure on education
Public expenditure on education	% of a country's GDP
Gross domestic product (GDP)	GDP in current prices in U.S. Dollars (millions)
GDP per capita	Price value in U.S. Dollars
Population	Mid-year population estimate (millions)
Primary education students	Number of students enrolled in primary education (thousands)
Secondary education students	Number of students enrolled in secondary education (thousands)
Tertiary education students	Number of students enrolled in tertiary education (thousands)

4 RESULTS

From figures 1 and 2, we see three clearly distinguishable clusters. Table 2 confirms the correlation between clusters and GDP per capita. we see that countries with a greater GDP per capita invest more public funding in public education and also exhibit greater school enrollment levels across primary, secondary, and tertiary education. More specifically, countries with GDP per capita less than 3500 U.S. dollars do not exhibit significant enrollment in tertiary education. Countries with GDP greater than 3500 U.S. Dollars

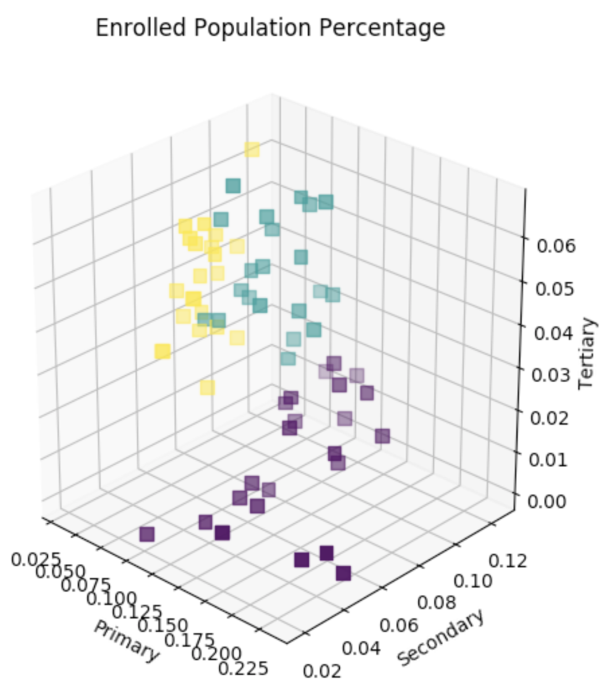

Figure 1

5 CONCLUSION

Our project goal was to analyze the relationship between a country's wealth, public expenditure on education, and recorded enrollment levels. We implemented clustering analysis to achieve our goal using a latent variable model and reached the following conclusions. Countries with a higher gross domestic product do appropriate more money towards public funding of education and in turn exhibit greater enrollment rates across primary, secondary, and tertiary education levels. We used Python, Tensorflow, Jupyter Notebook, Latex, and Overleaf for this project. Our code and related files are available on GitHub at https://github.com/SaumyashreeRay/PublicEducation.

6 FUTURE WORK

We used a limited number of features in our analysis. In the future, we would like to include more economic and education descriptors such as education enrollment distribution by gender or region. Furthermore, we would like to compare our latent variable model with other, simpler models such as k-means clustering to understand how they compare. We also could wish to extend the data by including information over a larger span of years.

7 ACKNOWLEDGEMENTS

We thank Professor Jorge Silva for teaching COMP 562: Machine Learning in this Fall 2019 semester and for encouraging us to explore a machine learning problem in greater depth.

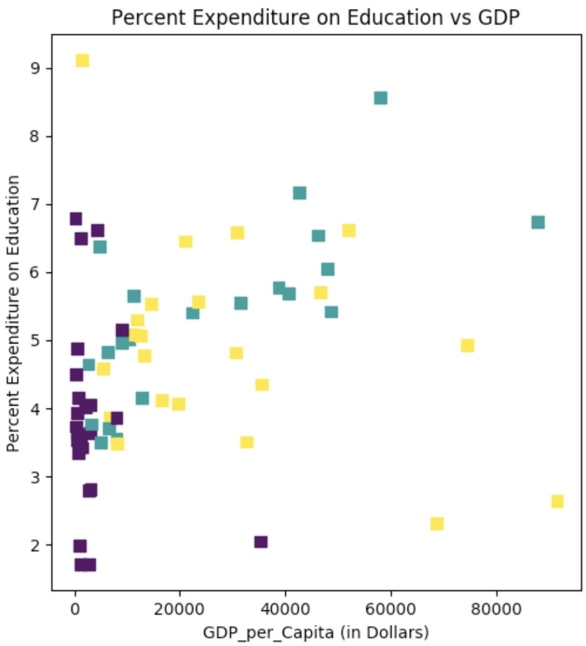

Figure 2

REFERENCES

[1] Patricia S. Abril and Robert Plant. 2007. Financing Higher Education in Africa. *Commun. ACM* 50, 1 (Jan. 2007), 36–44. https://doi.org/10.1145/1188913.1188915

[2] Mark C. Berger and Thomas Kostal. 2002. Financial Resources, Regulation, and Enrollment in US Public Higher Education. *Economics of Education Review* 21, 2 (2002), 101–110. https://doi.org/10.1016/S0272-7757(00)00065-0

[3] Andreas Bergh and Günther Fink. 2006. Higher Education: Does Public Expenditure Increase Enrollment? *Ratio Working Papers* 84 (2006). https://ideas.repec.org/p/hhs/ratioi/0084.html3

[4] David Greenaway and Michelle Haynes. 2003. Funding Higher Education in the UK: The Role of Fees and Loans. *The Economic Journal* 113, 485 (Feb. 2003), F150–F166. https://doi.org/10.1111/1468-0297.00102

[5] Eric A. Hanushek. 1989. The Impact of Differential Expenditures on School Performance. *Educational Researcher* 18, 4 (May 1989), 45–51. https://doi.org/10.3102/0013189X018004045

[6] Antonio Spilimbergo Karnit Flug and Erik Wachtenheim. 1998. Investment in Education: Do Economic Volatility and Credit Constraints Matter? *Journal of Development Economics* 55, 2 (1998), 465–481. https://doi.org/10.1016/S0304-3878(98)00045-5

[7] Amon Okapala and Comfort Okapala. 2006. The Effects of Public School Expenditure and Parental Education on Youth Literacy in Sub-Saharan Africa. *Journal of Third World Studies* 23, 2 (2006), 203–212. http://libproxy.lib.unc.edu/login?url=https://search.proquest.com/docview/233187820?accountid=14244

[8] Kathie Little Sandy Baum and Kathleen Payea. 2011. Trends in Community College Education: Enrollment, Prices, Student Aid, and Debt Levels. *The College Board: Trends in Higher Eductation* (2011). https://research.collegeboard.org/pdf/trends-2011-community-colleges-ed-enrollment-debt-brief.pdf

[9] Thomas M. Selden and Michael J. Wasylenko. 1995. Measuring the Distributional Effects of Public Education in Peru. *John Hopkins University Press for the World Bank* (1995). https://eric.ed.gov/?id=ED392578

[10] Kevin Sylwester. 2002. Can Education Expenditures Reduce Income Inequality? *Economics of Education Review* 21, 1 (Feb. 2002), 43–52. https://doi.org/10.1016/S0272-7757(00)00038-8