Books and Babies: An Analysis of the Relationship between Female Literacy Rate and Fertility Rate in South Asia (2000-2020)*

Jagpreet Singh

April 18, 2024

This paper utilizes data from the World Bank's Gender Statistics database to examine the relationship between the adult female literacy rate (% of females ages 15 and above) and fertility rate (total births per woman) in South Asia between 2000-2020. The analysis revealed a negative correlation between the two variables, indicating that as the literacy rate increased, the fertility rate decreased. These findings matter as they highlight the importance of investing in female education and empowerment to manage population growth and promote sustainable development. The insights can guide policymakers make decisions on education initiatives and reproductive health programs to further enhance socioeconomic progress in the region.

Introduction

According to the World Bank, the South Asian region comprises Afghanistan, Bangladesh, Bhutan, India, Maldives, Nepal, Pakistan, and Sri Lanka. It is the most populated region in the world with countries like India, Pakistan and Bangladesh which ranks among the top 10 in in terms of population size. Majority of these countries are also considered developing nations and play a major role in the world economy. In today's globalized world where population dynamics significantly influence a country's course it is necessary to analyze gender statistics indicators such as fertility rate and female literacy rate and how they relate to each other. By doing so we can gain valuable insights which can help make policy decisions on female education and empowerment that inturn help manage population growth and promote sustainable development.

*Code and data are available at: https://github.com/Jagpreet5ingh/female_fertility_rate_south_asia.git

In this paper, we will examine the relationship between the fertility rate (total births per woman) and adult female literacy rate (% of females ages 15 and above) in South Asia through a linear regression analysis. The estimand here is how fertility rate and female literacy rate are related. Specifically, we will focus on South Asia, the most populated region in the world. We will draw data from the World Bank website. Based on the analysis, we found that there is a negative relationship between fertility rate and literacy rate of adult females.

In section 1, we discuss the source of data used in this paper, methodologies that follow it, and data terminology. In section 2, we present the results of our analysis, focusing on the trajectory of fertility rate and female literacy rate over the last two decades in the region of South Asia. In section 3, we will analyze the trend by establishing a linear regression model. In section 4 we will present the result of the model in a graph. Finally, in section 5 we will further discuss the findings of the result.

Data

Data Description and Methodology

Table 1: A summary table of cleaned data

Year	fertility_rate_total	female_literacy_rate
2000	3.572224	44.92875
2001	3.515548	45.67490
2002	3.432268	46.99985
2003	3.339205	47.85153
2004	3.258046	48.89375
2005	3.166954	48.09562
2006	3.073746	48.75041
2007	2.999592	49.91730
2008	2.929633	50.99176
2009	2.877746	52.00314
2010	2.807602	55.80537
2011	2.742788	56.80440
2012	2.678348	57.57100
2013	2.621332	58.56601
2014	2.530008	59.59675
2015	2.498366	60.73805
2016	2.475185	62.34228
2017	2.406131	63.23561
2018	2.379341	63.78237
2019	2.320116	63.74677

Year	fertility_rate_total	female_literacy_rate
2020	2.267098	64.60739

Data Visualization

Trend of Female Literacy Rate in South Asia (2000-2020)

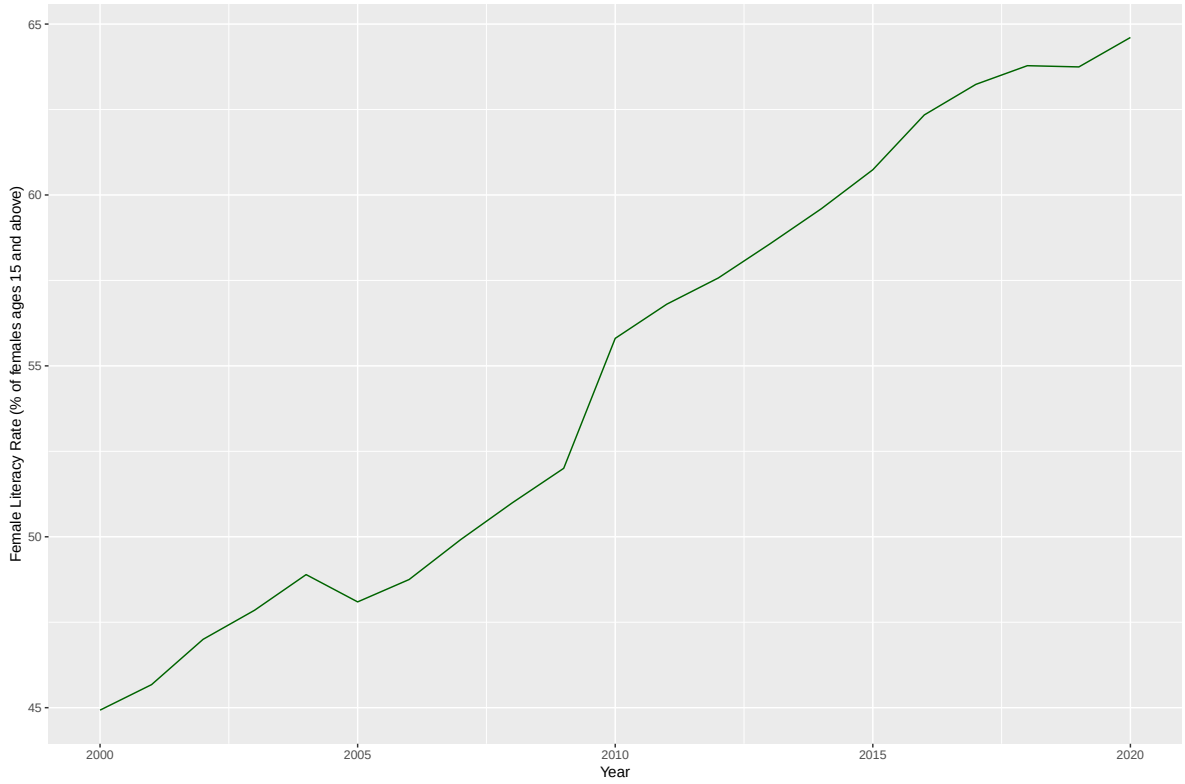

Figure 1: literacy rate, adult females (% of females ages 15 and above) in South Asia (2000-2020)

Trend of Fertility Rate in South Asia (2000-2020)

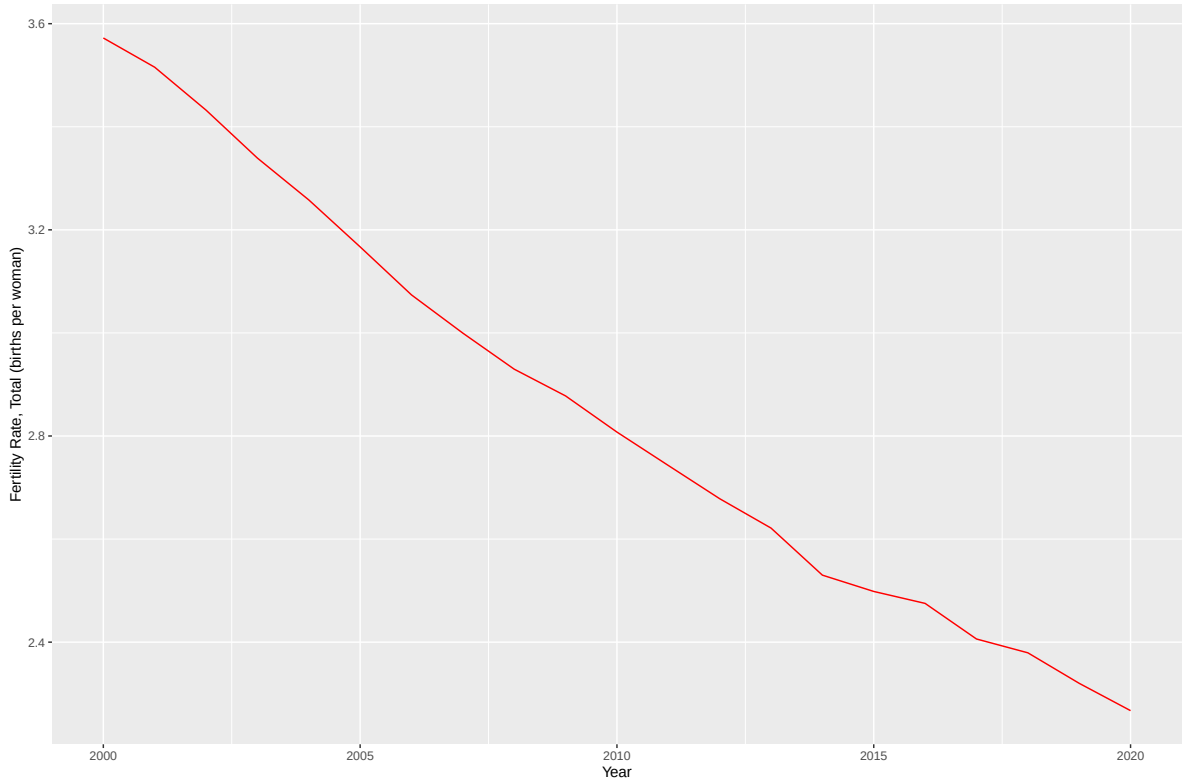

Figure 2: literacy rate, Fertility Rate, Total (births per woman) in South Asia (2000-2020)

Variables of Interest

Model

Based on the preliminary analysis, we observed a negative relationship between fertility rate, total (births per woman) and literacy rate, adult females (% of females ages 15 and above). This indicates a potential linear regression.

Below is the equation of the linear regression model:

$$Yij = \beta 0 + \beta 1 Xij + \varepsilon ij$$

where:

- Y is the Fertility Rate
- X is the Female Literacy Rate
- Y_{ij} is the Fertility Rate for observation j in year i.
- X_{ij} is the Female Literacy Rate for observation j in year i.
- β_0 is the intercept/constant term, which represents the expected value of Fertility Rate when the Female Literacy Rate is equal to zero.
- β_1 is the slope coefficient or the estimated change in Fertility Rate for a one-unit increase in the Female Literacy Rat.
- ϵ_{ij} is the error term or the deviation of the actual value of Fertility Rate from the predicted value based on the regression equation.

This linear regression model aims to estimate the values of 0 and 1 such that the model fits the data well, and predicts the expected value of the Fertility Rate for different values of the Female Literacy Rate. The statistical significance of 1 can be assessed using a t-test, which tests whether the estimated coefficients are significantly different from zero. If the p-value of the t-test is less than the selected level of significance, we can conclude that there is a significant relationship between the Female Literacy Rate and the Fertility Rate.

Results

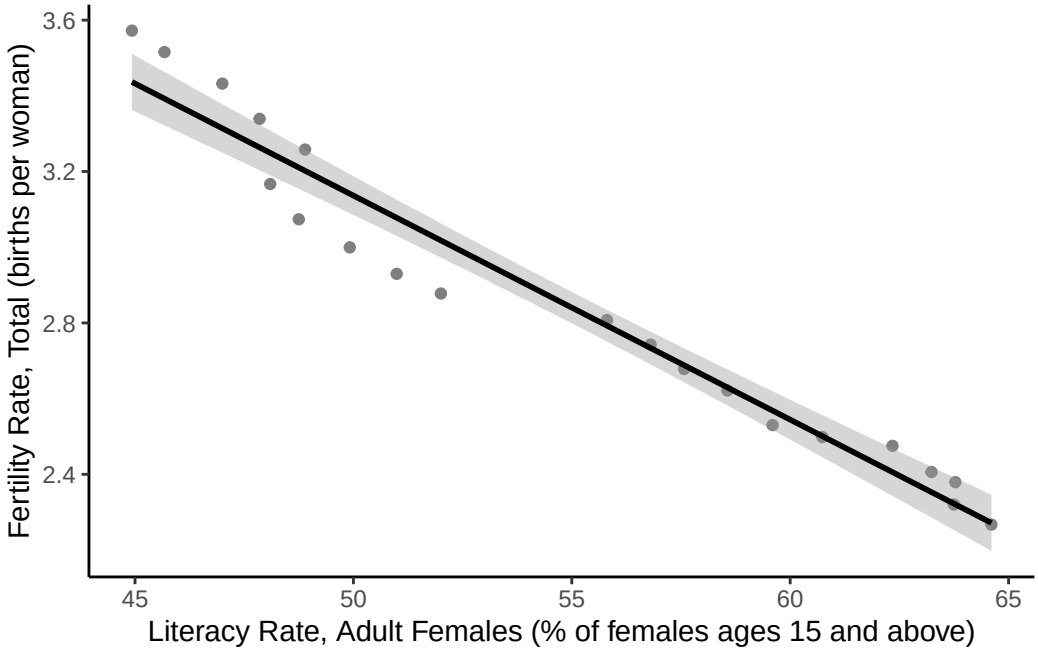

Figure 3: The model shows the relationship between Female Literacy Rate and Fertility Rate in South Asia (2000-2020)

Table 2: T-Value and P-Value of Regression

	t_value	p_value
(Intercept)	36.95598	0
female_literacy_rate	-19.81133	0

Model Justification

Discussion