**Project – Assignment 5:**

You will be writing up your data and methodology and results for the project in this assignment.  Please address each of the following in a typed document.  Have each point be a heading in your paper.  The headings can be numbered (1) through (9).

1. **State your research question**

How has the Covid-19 pandemic impacted primary school enrollment in low- and middle-income countries?

1. **State your hypothesis**

Primary school enrollment in low- and middle-income countries significantly decreased during the COVID-19 pandemic compared to pre-pandemic levels.

1. **Describe your dependent variable(s) including source (make sure it is a complete reference in APA style not just a web address), how it is measured, the year(s) the variable is from, and why this is an appropriate variable to test your hypothesis.  (Don’t forget real, current or PPP in your discussion)**

**Variable:** Primary School Enrollment Rate (Net, % of primary school age children)   
Dependent Variable of the Data Set

**Source Citation:**   
World Bank. (2023). *Primary school enrollment, net (% of primary school age children)*. World Development Indicators. <https://data.worldbank.org/indicator/SE.PRM.NENR>

**Measurement:**   
This variable measures the net enrollment rate (NER) for primary education, which is defined as the percentage of children of the official primary school age who are enrolled in primary school. It reflects the extent to which children in the correct age group are participating in the primary education system.

**Years Covered:**   
From the study source, we used data from 2019 to 2022, allowing us to analyze the pre- and post-Covid-19 comparison to evaluate the impact of the pandemic on lower- and middle-income counties.

**Why the Variable Is Appropriate for Our Hypothesis:**   
The net enrollment rate is a direct indicator of participation in the primary education system. It is especially relevant for low- and middle-income countries, where educational disruptions during the Covid-19 pandemic were most pronounced due to limited remote learning infrastructure and economic challenges that may have taken place within the countries. Using this variable enables our group to analyze how effectively these countries were able to maintain or recover primary school participation following the pandemic.

1. **Describe each one of your independent variables, including the source (make sure it is a complete reference in APA style, not just a web address), how it is measured, the year(s) the variable is from, what the predicted relationship with your dependent variable is, and why you have included it. (Don’t forget real, current or PPP in your discussion and the form of the variable—is it ln(X), for example)  If any of the variables have changed from your submission in assignment 4, then please resubmit a correlation coefficient table.**

**Variable:** Education Expenditure (% of GDP)

**Source Citation:**   
World Bank. (2023). *Government expenditure on education (total (% of government expenditure)*. World Development Indicators. <https://data.worldbank.org/indicator/SE.XPD.TOTL.GB.ZS>

**Measurement:**   
This variable is measured as a percentage of the total government budget spent on education. This variable captures national prioritization of educational investment in each country measured.

**Years of Variable:** 2019 to 2022

**Predicted Relationship with the Dependent Variable:**   
We expect a positive relationship with the dependent variable; as education spending rises in low- and middle-income countries, enrollment in primary school will be expected to increase due to improved school resources, teacher availability, and infrastructure within those countries.

Variable: ln(GDP per Capita) (Current constant US$)

**Source Citation:**   
World Bank. (2023). Ln(GDP per capita) (current US$). *World Development Indicators*. <https://data.worldbank.org/indicator/NY.GDP.PCAP.CD>

**Measurement:**   
This is how much income, on average, each person in a country earns, based on total GDP divided by the population. It’s in current U.S. dollars, not adjusted for inflation.

**Years of Variable:** 2019 to 2022

**Predicted Relationship with the Dependent Variable:**   
I think this will also have a positive relationship with enrollment. Countries with higher income levels can afford better education systems, and families are more likely to send their kids to school when they don’t have to worry as much about basic survival needs or putting their kids to work.

Variable: Year (Dummy Variable)

**Source Citation:**   
World Bank. (2023). World Development Indicators. *World Bank Group*. <https://databank.worldbank.org/source/world-development-indicators>

**Measurement:**   
Year is used here as a dummy variable, and we are using 2018 more specifically as our dummy. We did this to capture any trends or changes over time from 2019 to 2022 that could be affecting enrollment, regardless of the country. Each year gets coded as its own value (like 0/1), which helps account for any outside factors like global events or policy shifts that could impact education across the board.

**Years of Variable:** 2019 to 2022

**Predicted Relationship with the Dependent Variable:**   
This one is less about predicting a specific direction and more about capturing overall time trends. We included it to help control for year-by-year changes that might influence enrollment even if the other variables stay the same, but we expect the dependent variable to decrease as the years go on. We thought that the year changing would decrease the primary enrollment because of Covid and there would be lasting effects.

1. Describe your sample including the countries and years in your dataset.  Include a discussion of how you choose your sample.  (Is the sample a global sample, area specific, income specific, etc.?)

**Sample Description**

Our dataset includes 102 countries and spans the years 2019 to 2022. The countries in the sample are all classified as low- and middle-income economies, according to World Bank classifications. We intentionally chose to focus on this income group because primary school enrollment and education-related investments can vary a lot depending on a country’s development level. By limiting the sample to low- and middle-income countries, We were able to better understand the specific factors influencing education access in places where resources might be more limited or unevenly distributed.

This is an income-specific sample rather than a global or regional one. It covers a broad range of countries across different continents, including Albania, Algeria, Bangladesh, Belize, and many others. These countries were included based on available data for the variables we selected, such as education spending, ln(GDP per capita), and access to clean water.

The dataset includes observations from five consecutive years, 2019 through 2022,which gives us a chance to look at trends over time while keeping the focus on recent data. We also included "Year" as a variable to help account for time-based effects in our analysis.

1. Submit your complete dataset.
2. Create a table of your results in word.  Here is an example of what it should look like.

|  |  |
| --- | --- |
|  | Primary Enrollment (% Gross) |
| Constant | 87.71\*\*\* |
| (6.36) |
| ln(GDP Per Capita) | 0.53 |
| (0.71) |
| Education Expenditure      Y2 (2019)      Y3 (2020)      Y4 (2021)      Y5 (2022) | 0.68\*\*\* |
| (0.12)    -.0188  (1.9)    -0.533  (1.9)    -1.27  (1.89)    -.022  (2.08) |
|  |  |
|  |  |
| Observations | 384 |
| F statistics | 5.47 |
| R2 | 0.08 |

Note: Standard errors for independent variables are shown in parentheses. The symbols \*,\*\*,\*\*\* correspond to a 10% , 5% and 1% level of significance.

1. **Discuss your results.  Make sure to interpret the coefficient of all of your variables.  Explain whether your hypothesis was confirmed.**

For every 1% increase in ln(GDP per capita), the gross enrollment rate decreases by 0.53 percentage points. This is not significant, suggesting that ln(GDP per capita) does not have a meaningful effect on primary school enrollment in terms of the effects of COVID-19. In 2019 the coefficient was –1.382, showing enrollment was 1.38 points lower than in 2018, which is not significant (0.779). In 2020 the coefficient was –1.583, showing enrollment declined 1.58 points, not significant (0.722). In 2021 the coefficient was -2.220, showing a further decline in primary school enrollment, which is also not significant (0.638). Finally, in 2022, the coefficient was -6.227, which shows the largest drop in primary school enrollment but is not significant (0.118) in reference to primary school enrollment.

Overall, none of the individual years were statistically significant. This suggests that although the pandemic years are associated with lower enrollment, we can’t say with confidence that the decline was caused by the specific factors modeled here. The regression model as a whole explains only a small portion of the variation in enrollment (R² = 0.08), indicating that other important variables like internet access, rural vs. urban schooling, or health policy, or any other variable that can relate to the dependent variable, may need to be included in future models to better understand what affects primary school enrollment across countries during this time period.

1. **Discuss one shortcoming you see with your project.**

One of the shortcomings was that we needed to use one less variable in our data, so this made it harder for us to get an accurate regression. In addition, it is hard to predict primary school enrollment as a dependent variable because of the lack of consistency in the data that we had to predict. If we were to do it over again, we would have a lot more variables and choose some better ones, too. We have a low R2 score, which means that we only accounted for 8% of the variation in primary enrollment. Lastly, we had uneven data coverage when it came to the countries. Some of the countries didn’t have data, so they had to get dropped for the year. This led to fewer observations, making the regression less accurate.  Another shortcoming could be how we used a scale of national level in low

1. **Create a table in word that includes each of your variables, including your dummy variable(s), and report the mean, standard deviation, minimum value, and maximum value in your sample.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Mean** | **Stand Dev** | **Min** | **Max** |
| **GDP Per Capita**    **Education Expenditure**    **Primary Enrollment**    **Y1 (2019)**    **Y2 (2020)**    **Y3 (2021)**    **Y4 (2022)** | 4008.599      15.3898      101.4549 | 3221.849      5.2789      14.3318 | 253.6916      2.566      8.4479 | 16242.06      35.0058      156.8019 |