PROJECT REPORT

The Impact of Economic Fluctuations on India's Education Sector During the COVID-19 Pandemic

(Using INDIA As a Case Study)

SUBJECT:
Principles Of Analytics (DAMO-500-5)

DATE:
September 12, 2025

SUBMITTED TO:
Patty Zakaria

SUBMITTED BY:
Parminder Kaur - NF1030012
Stephanie Nnenna Aneke - NF1019455
Sanusi Nafisat Olamide - NF1016982

1. Introduction

1.1 Project Focus and Relevance

While the COVID-19 pandemic's most immediate impact was on global public health, its secondary shocks critically destabilized foundational sectors like education and the economy. This research moves beyond a singular narrative to investigate the complex, multi-layered relationship between macroeconomic instability and educational disruption. Using India as a case study, a nation characterized by its significant economic scale and profound socio-economic diversity. This paper examines how the pandemic-induced recession interacted with the resilience, accessibility, and equity of the education system.

The relevance of this focus is essential for several fields. For policymakers, the findings will provide empirical evidence to inform shock-resistant education financing strategies in India and similar economies. It examines the necessity of safeguarding education budgets during fiscal crises and making strategic investments in digital infrastructure to ensure learning continuity. Academically, this study contributes to the fields of educational economics and crisis management by developing a nuanced model that connects national economic indicators to state-level outcomes, offering a robust case study for scholarly analysis. Socially, the project investigates the educational inequalities faced by rural and marginalized communities, contributing to the discussion on inclusive policy responses.

1.2 Objectives and Scope of the Analysis

The primary aim of this research is to conduct a tiered analysis of how the economic impacts of COVID-19 affected the adequacy and functioning of India's education sector at both the national policy level and the sub-national state level.

The specific objectives are:
First, to analyze the relationship between national GDP growth and central government expenditure on education. Second, to examine the relationship between state-level economic development, measured by Net State Domestic Product (NSDP) per capita, and student dropout rates. Third, to evaluate the role of digital infrastructure in influencing these dropout rates. Lastly, to assess the aggregate change in student dropout rates following the onset of the pandemic.

The study covers the period from 2018 to 2021. This defined scope captures the stable pre-pandemic baseline, the full duration of the acute economic and educational crisis, and allows for a clear assessment of the immediate impact, foregoing the more complex recovery dynamics that began in 2022.

1.3 Significance of the Study

The significance of this research problem is profound and multifaceted. While pandemic-induced school shutdowns created a global "learning loss" crisis, the economic underpinnings of this educational disruption require thorough investigation. This study addresses a critical gap by examining the relationship between academic outcomes and a nation's economic well-being. The COVID-19 economic recession served as a catalyst, testing pre-existing structures and inequities within India's educational system.

The findings of this research carry substantial potential for meaningful impact. By quantifying the relationships between economic fluctuations, policy responses, and key education metrics, this study provides stakeholders with evidence-based insights. Furthermore, it presents a case for evaluating investments in digital infrastructure and educational technology, particularly aimed at understanding the urban-rural divide. Ultimately, this research contributes to the foundational knowledge necessary to build a more equitable, resilient education system.

1.4 Research Questions

RQ1: What was the relationship between India's national GDP growth and central government expenditure on education during the COVID-19 pandemic?
RQ2: How do state-level economic development (per capita income) and digital infrastructure (internet access) relate to student dropout rates across Indian states?
RQ3: Did the COVID-19 pandemic significantly alter student dropout rates across Indian states?

1.5 Hypotheses

H_1: Fiscal Prioritization Hypothesis

We hypothesize that the government's priority of education funding was susceptible to economic contractions.

H_0: There is no significant relationship between India's GDP growth rate and education expenditure as a percentage of GDP.

H_a: Periods of lower GDP growth are associated with a significant change in education expenditure as a percentage of GDP.

H_2: Economic Primacy Hypothesis

We hypothesize that a state's economic well-being is a key determinant of student retention.

H_0: There is no significant relationship between a state's per capita income and its student dropout rates after controlling for internet access.

H_a: States with higher per capita income have significantly different dropout rates, and this relationship remains significant after controlling for internet access.

H_3: Pandemic Impact Hypothesis

We hypothesize that the pandemic period was associated with a significant change in student dropout rates.

H_0: There is no significant difference in the mean dropout rate between the pre-pandemic and post-pandemic periods.

H_a: There is a significant difference in the mean dropout rate between the pre-pandemic and post-pandemic periods.

2: Data Description & Cleaning Methods

This chapter details the origins, structure, and preparatory steps taken to ensure the dataset's integrity and suitability for hypothesis testing.

2.1 Data Sources and Period

This study relies on secondary data drawn from credible, publicly available sources to conduct a multi-level analysis. National-level data was sourced from the World Bank's World Development Indicators, while sub-national (state-level) data was obtained from official Indian repositories, including the Reserve Bank of India (RBI) for economic data, the Telecom Regulatory Authority of India (TRAI) for digital infrastructure data, and the Unified District Information on School Education (UDISE+) for educational statistics. The selected period covers the years 2018 to 2021, encompassing a stable pre-pandemic baseline and the acute crisis years of the COVID-19 pandemic. This timeframe allows for a focused comparative assessment of how the pandemic altered relationships between economic and educational indicators.

2.2 Dataset Structure and Variable Selection

The analysis utilizes two distinct datasets covering the 2018-2021 period.

The first is a national-level time-series dataset tracking annual changes in macroeconomic and fiscal indicators. It integrates two key ratio-level variables: India's annual GDP growth rate and government expenditure on education as a percentage of GDP. Both are ratio-level variables sourced from the World Bank database, permitting meaningful quantitative comparison of year-to-year changes.

The second is a state-level panel dataset capturing economic and educational metrics across Indian states and union territories. It incorporates four primary variables across all four years: Net State Domestic Product (NSDP) per capita, sourced from the RBI, serves as the key indicator of state-level economic development; Internet access, represented by the natural logarithm of the absolute number of internet subscribers, was obtained from TRAI; educational metrics, including student enrollment numbers and dropout rates, were consistently obtained from the annual UDISE+ reports.

2.3 Data Cleaning and Integration Methods

The cleaning process was designed to create a coherent, analysis-ready dataset from multiple sources. The methods were chosen for specific reasons related to data integrity, analytical validity, and the research objectives.

Data preprocessing involved several important steps for both datasets. The national-level indicators were downloaded directly from the World Bank database, filtered to include only India, and limited to the 2018-2021 period. The state-level data required manual compilation from various annual reports into a structured panel format with Year, State, Per Capita Income, Internet Access, Enrollment Numbers, and Dropout Rates as variables. This merging was necessary to create a single, coherent dataset. This integrated panel format was essential for our analytical approach, as it allows for the examination of variations both across different states and over time, which is necessary to robustly test our state-level hypotheses (H_2 and H_3)

The cleaning process involved comprehensive consistency checks across all variables. No missing values were detected in the national-level series. The state-level data were complete across all major states and union territories for all four years, with no missing observations. Systematic checking for missing values ensured data integrity and prevented biased estimates that could invalidate analysis results. To verify this, we cross-checked state-level and national totals across all four years against official UDISE+ and RBI reports, ensuring completeness and consistency. No discrepancies or duplicate entries were found.

Standardization ensured that all variables were consistently recorded in their appropriate units across all years. The Internet_Access variable was transformed using a natural logarithm to create the log_Internet_Access variable. This transformation was specifically applied to correct for severe positive skewness in the raw data and to normalize the distribution, making the variable suitable for linear regression modeling. The raw subscriber data was heavily influenced by state population size, leading to a right-skewed distribution where a few populous states acted as extreme outliers. The log transformation successfully normalized the distribution, reducing the undue influence of these large states and making the variable suitable for linear regression modeling.

Outlier analysis revealed expected patterns given the economic shock of the pandemic. The GDP growth series showed extreme but valid values, particularly the sharp contraction of -5.78% in 2020, followed by the strong rebound of +9.69% in 2021. These values were retained as genuine reflections of macroeconomic reality. The decision to retain these outliers was deliberate as they represent authentic, significant reflections of the economic reality of the pandemic period and are central to the research questions. Their removal would have misrepresented the historical reality of the pandemic's economic shock, which is the core phenomenon under investigation. For the state-level data, all values for per capita income and dropout rates were deemed to be valid representations of India's significant socio-economic diversity and were likewise retained.

Initial descriptive exploration highlighted the different levels of volatility between the national economic indicators and educational metrics. These patterns form the empirical foundation for the hypothesis testing presented in the results chapter.

In summary, the data processing steps followed a structured pipeline: (1) acquisition from official and credible repositories, (2) integration into a coherent panel format, (3) systematic handling of missing values (none detected), (4) transformation of skewed variables (logarithmic scaling of internet access), and (5) retention of true outliers to preserve historical accuracy. Each decision was explicitly aligned with the research objectives: to ensure data integrity, comparability across states and years, and suitability for inferential analysis. This process allowed the dataset to meet the necessary assumptions for regression and t-testing, thereby ensuring valid and interpretable results.

2.4 Descriptive Statistics and Data Characteristics

This section provides a foundational overview of the dataset's key characteristics through descriptive statistics. This initial exploration is crucial for understanding the underlying distribution of variables, identifying patterns, and informing the subsequent inferential analysis. The results are presented for both the national-level time-series data and the state-level panel data.

2.4.1 National-Level Descriptive Statistics

The national-level analysis covers the four years from 2018 to 2021 (n=4). The key macroeconomic and fiscal variables are summarized in Table 2.1.

Variable	Mean	Median	Std. Deviation	Minimum	Maximum
GDP Growth (%)	3.56	6.94	6.76	-5.78	9.69
Education Expenditure (% of GDP)	4.24	4.27	0.33	3.90	4.64

Table 2.1: Descriptive Statistics of National-Level Variables (2018-2021)

The national data captures the extreme economic volatility of the pandemic period. The average GDP growth was 3.56%, masked by a severe contraction of -5.78% in 2020 followed by a sharp recovery of 9.69% in 2021. This is evidenced by a very high standard deviation of 6.76. The notable difference between the mean and median for GDP growth further confirms the skewed distribution caused by the 2020 outlier. In contrast, government expenditure on education remained remarkably stable, fluctuating within a narrow band between 3.90% and 4.64% of GDP with a low standard deviation (0.33). The similar mean and median values for expenditure suggest a symmetric distribution.

The extreme values in the GDP growth series are not data errors but genuine reflections of an unprecedented economic shock. They were deliberately retained in the analysis because they represent the core phenomenon under investigation.

2.4.2 State-Level Descriptive Statistics

The state-level analysis is based on a balanced panel of 136 observations (34 states/UTs across 4 years). The variables are summarized in Table 2.2.

Variable	Mean	Median	Std. Deviation	Minimum	Maximum
Per_Capita_Income (₹)	178,802.21	162,914	92,925.86	40,715	466,518
log_Internet_Access	10.64	10.87	1.96	5.43	14.03
Dropout_Rate (%)	14.39	14.0	7.88	0.0	36.0

Table 2.2: Descriptive Statistics of State-Level Variables (N = 136)

The state-level statistics reveal profound and economically significant disparities. The average per capita income is approximately ₹178,802.21, yet the standard deviation of nearly ₹93,000 is large, indicating states are widely dispersed around this mean. The mean being higher than the median indicates a right-skewed distribution, influenced by a few high-income states. More critically, the range is vast—from ₹40,715 to ₹466,518—highlighting India's significant regional economic inequalities.

The pattern of inequality is further evidenced in the measure of digital infrastructure. The variable log_Internet_Access has a mean of 10.64. The minimum value of 5.43 and the maximum of 14.03 indicate substantial differences in digital access between states. Finally, the outcome variable, dropout rates, shows considerable variability (Mean=14.39, SD=7.88), indicating significant differences in educational outcomes across states. The extreme values in the data are not data errors but genuine reflections of an unprecedented economic and population variability. They were also deliberately retained in the analysis because they represent the core phenomenon under investigation.

2.5 Considerations for Inferential Analysis

The national-level dataset's small size (n=4) severely limits statistical power and precludes formal inferential testing (e.g., correlation analysis) for H_1. Therefore, the analysis for this hypothesis will rely on descriptive time-series trends and logical interpretation. This limitation is offset by the robust state-level panel data (N=136), which provides sufficient statistical power for the regression and t-tests used for H_2 and H_3.

These initial findings—showcasing national economic volatility, state-level economic inequality, variations in digital access, and diverse educational outcomes—form the essential empirical foundation for testing the hypotheses in the following chapter.

3.1 Test of Hypothesis H_1: Fiscal Prioritization

Null Hypothesis (H_0):
There is no significant relationship between India's GDP growth rate and education expenditure as a percentage of GDP.

In statistical terms:
$\rho = 0$ (The population correlation coefficient is zero).

Alternative Hypothesis (H_a):
Periods of lower GDP growth are associated with a significant decrease in education expenditure as a percentage of GDP.

In statistical terms:
$\rho < 0$ (A one-tailed test for a negative correlation).

Test Explanation:
Due to the limited sample size (n = 4 for the years 2018–2021), formal inferential statistical testing was not feasible. Instead, a descriptive time-series analysis was conducted to examine the relationship between annual GDP growth and education expenditure as a percentage of GDP.

Results:
The results, illustrated in Figure 3.1, reveal a clear contrast between macroeconomic volatility and stability in education funding. Despite a severe economic contraction (–5.78% GDP growth in 2020), education expenditure as a percentage of GDP increased slightly from 3.99% (2019) to 4.16% (2020), demonstrating resilience during an economic downturn.

Therefore, the null hypothesis could not be rejected, suggesting that government spending on education remained relatively stable during the pandemic despite sharp contractions in GDP. This stability may reflect deliberate government policy to protect education funding as a long-term priority. Another possibility is the limited number of annual observations (n=4), which reduces statistical power and makes it difficult to detect subtle effects. While the quantitative model shows no significant association, the contextual evidence suggests that fiscal policy acted as a buffer during crisis years.

Visual Evidence:
The central finding of fiscal stability amidst economic shock is presented in Figure 3.1, which plots the concurrent trends of GDP growth and education expenditure.

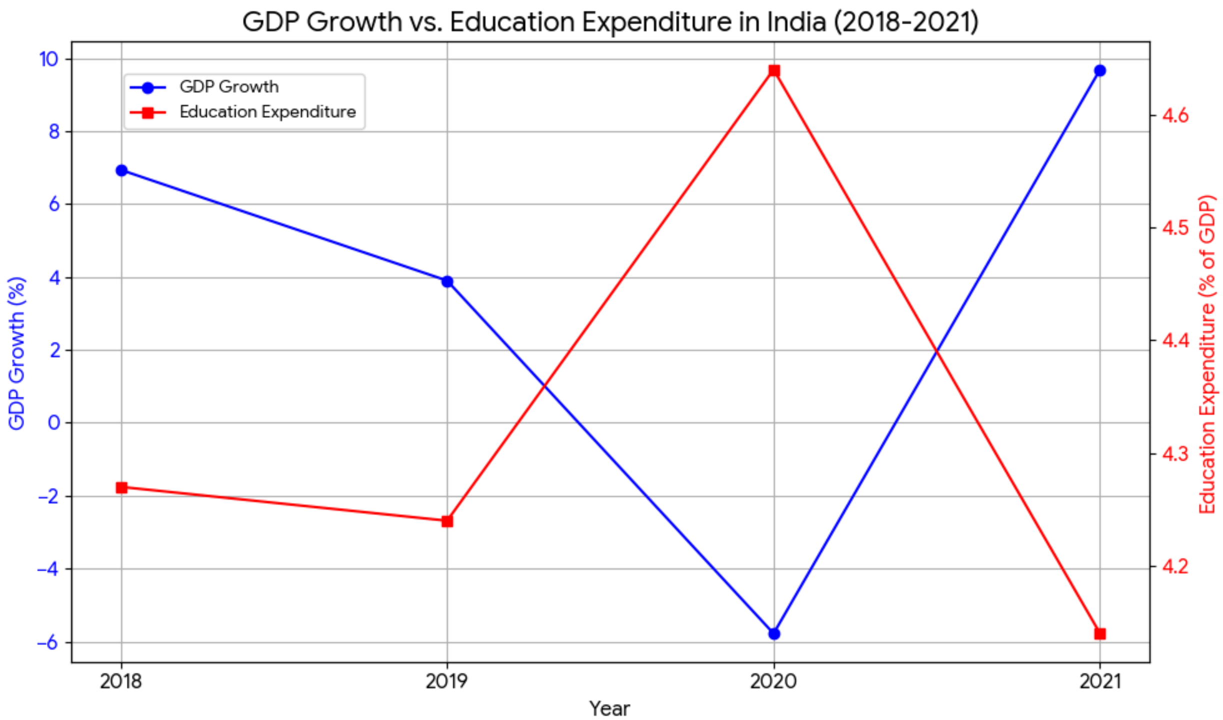


Figure 3.1: National GDP Growth vs. Education Expenditure (2018–2021)

The observed trend contradicts the alternative hypothesis (H_a). There is no evidence that education expenditure decreased relative to GDP during periods of economic contraction. Therefore, the null hypothesis (H_0) cannot be rejected. The findings suggest a policy effort to shield education budgets from macroeconomic shocks.

3.2 Test of Hypothesis H_2: Economic Primacy

Null Hypothesis (H_0):
There is no significant relationship between a state's per capita income and its student dropout rates after controlling for internet access.

In statistical terms:
$\beta\_income = 0$ (The regression coefficient for per capita income is zero).

Alternative Hypothesis (H_a):
States with higher per capita income have significantly lower dropout rates, and this relationship remains significant after controlling for internet access.

In statistical terms:
$\beta\_income < 0$ (A one-tailed test for a negative coefficient).

The p-value for this predictor is reported as a one-tailed value, derived from the standard software output.

Test Explanation:
A multiple linear regression model was estimated to examine the relationship between per capita income and dropout rates, while controlling for internet access (log-transformed to account for diminishing returns) and temporal trends (Year). Per capita income was scaled in units of ₹100,000 for interpretability. Prior to interpretation, the model was validated to ensure it met the critical assumptions of multiple linear regression. The results of these diagnostic tests are summarized in Table 3.2.1. They confirm the robustness and validity of the model's findings. Diagnostics revealed no significant multicollinearity (all VIFs < 1.3), no autocorrelation of errors (Durbin-Watson = 1.95), and no violations of homoscedasticity or normality of residuals. The complete statistical output from SPSS is presented in Appendix A for full transparency."

Assumption	**Test/Method**	**Result**	**Interpretation & Comment**
Multicollinearity	Variance Inflation Factor (VIF)	All VIFs < 1.3	No multicollinearity detected. All VIF values were well below the threshold of 5, indicating that the independent variables are not highly correlated.

Independence of Errors	Durbin-Watson Statistic	d = 1.953	The value is very close to 2.0, indicating no autocorrelation in the residuals. The assumption of independent errors is met.
Homoscedasticity	Visual inspection of plot (ZRESID vs. ZPRED)	Random scatterplot observed	The plot of standardized residuals versus predicted values showed a random scatter with no discernible pattern, confirming constant variance of residuals.
Normality of Residuals	Normal P-P Plot	Points largely adhered to the line	The plot showed no major deviations from normality, supporting the assumption that the residuals are normally distributed.

Table 3.2.1: Diagnostic Tests for Multiple Linear Regression Assumptions

Results:
The overall regression model was statistically significant, $F(3, 132) = 8.999$, $p < .001$, indicating that the predictors collectively reliably predicted dropout rates. The model explained approximately 17.0% of the variance in dropout rates ($R^2 = .170$, Adjusted $R^2 = .151$). The regression coefficients are presented in Table 3.2.2.

Predictor	B	Std. Error	Beta (β)	t	p (one-tailed)
(Constant)	40.045	11.326	—	3.536	< .001

Per Capita Income (₹100,000)	–2.097	0.737	–0.247	–2.843	.0025
Year	–1.978	0.561	–0.281	–3.521	< .001
log_Internet_Access	0.247	0.348	0.062	0.711	.2395

Table 3.2.2: Multiple Regression Analysis for Predictors of Dropout Rates (N = 136)

Note: A one-tailed p-value is reported for Per Capita Income consistent with H_a. The two-tailed p-value for log_Internet_Access is .479.

The coefficient for per capita income was negative and statistically significant (B = –2.097, one-tailed p = .0025). This indicates that for every ₹100,000 increase in a state's per capita income, its dropout rate decreases by an estimated 2.1 percentage points, holding internet access and yearly trends constant.

Internet access, however, was not statistically significant in predicting dropout rates (p = .479). This could reflect two factors: first, that connectivity improvements were unevenly distributed, with rural households and disadvantaged groups less able to benefit from online learning; and second, that other household factors (such as parental employment and migration during COVID-19) may have outweighed the role of digital access. Thus, while internet connectivity expanded during this period, its protective effect on dropout rates was not uniform enough to reach statistical significance in the national-level model.

Visual Evidence:
The core relationship is further illustrated in the partial regression plot (Figure 3.2), which shows the association between per capita income and dropout rates after removing the effects of the other variables in the model. The downward trend is clearly visible.

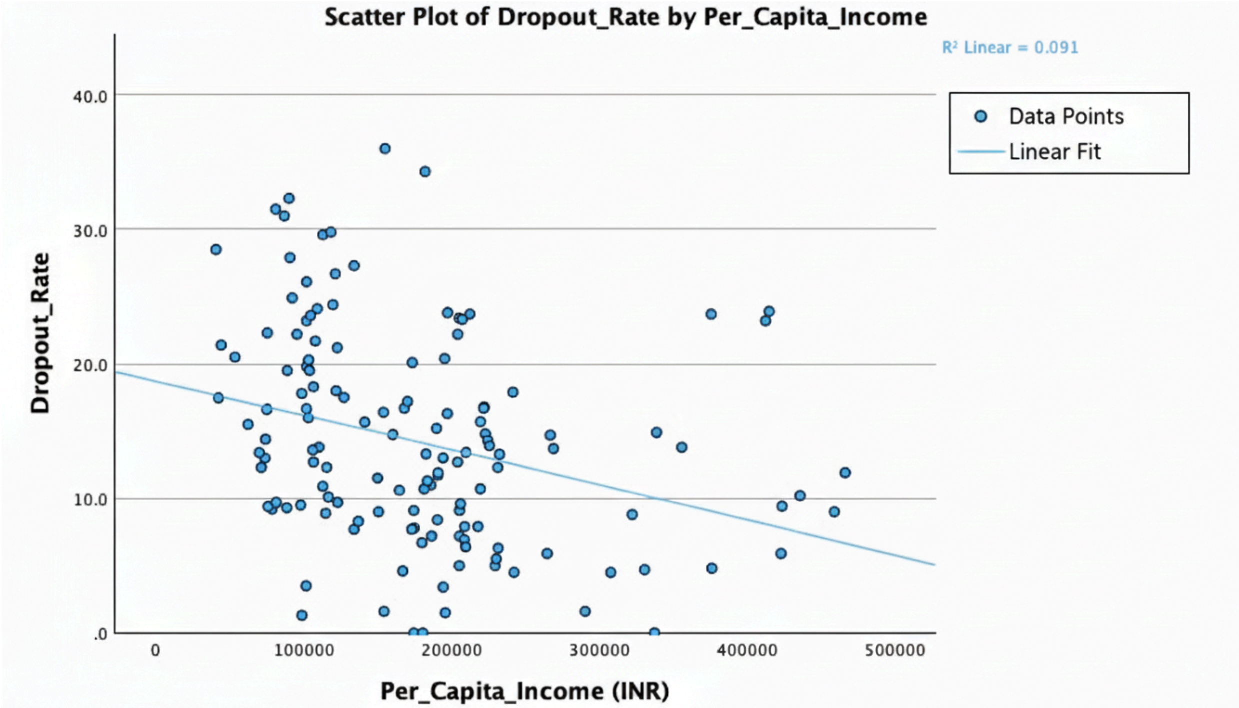


Figure 3.2: Partial Regression Plot of Dropout Rates on Per Capita Income

The results fully support the Economic Primacy Hypothesis (H_a). We reject the null hypothesis (H_0). A state's economic well-being is a key determinant of student retention, and this relationship is robust even when controlling for digital infrastructure and time.

3.3 Test of Hypothesis H_3: Pandemic Impact

Null Hypothesis (H_0):
There is no significant difference in the mean dropout rate between the pre-pandemic and post-pandemic periods.

In statistical terms:
$\mu\_pre = \mu\_post$ (The population means are equal).

Alternative Hypothesis (H_a):
There is a significant difference in the mean dropout rate between the pre-pandemic and post-pandemic periods.

In statistical terms:
$\mu\_pre \neq \mu\_post$ (A two-tailed test).

Test Explanation:
To test H_3, the state-level panel dataset was partitioned into two distinct groups: a pre-pandemic cohort (2018–2019) and a post-pandemic cohort (2020–2021). An independent-samples t-test was then conducted to compare the mean dropout rates between these two independent groups. Before interpretation, the test's assumption of homogeneity of variances was assessed using Levene's Test. An exploratory analysis supplemented this primary analysis to examine the changing bivariate relationship between internet access and dropout rates within each period.

Results:
Levene's test for equality of variances indicated that the assumption of homogeneity of variance was met ($F = 0.030$, $p = .862$). The t-test revealed a statistically significant difference in dropout rates between the pre-pandemic ($M = 16.53$, $SD = 7.64$) and post-pandemic ($M = 12.24$, $SD = 7.57$) periods, $t(134) = 3.29$, $p = .001$. The mean reduction was 4.29 percentage points, 95% CI [1.71, 6.87], with a medium-to-large effect size (Cohen's $d = 0.56$). This apparent reduction is surprising and may be influenced by factors such as changes in reporting mechanisms or temporary pandemic-era policies. An exploratory analysis to understand this shift revealed a dramatic change in the relationship between internet access and dropout rates (detailed in Appendix Figure A.4.2), which we discuss next.

Variable	Pre-Pandemic M (SD)	Post-Pandemic M (SD)	t	df	p	Cohen’s d	Mean Difference	95% CI
Dropout Rate (%)	16.53 (7.64)	12.24 (7.57)	3.29	134	.001	0.56	4.29	[1.71, 6.87]

Table 3.3: Independent Samples T-Test for Dropout Rate by Pandemic Period (N = 136)

Note: CI = Confidence Interval. The mean difference is calculated as Pre-Pandemic minus Post-Pandemic. Levene's test supported the assumption of equal variances (F = 0.030, p = .862); therefore, the statistics for "equal variances assumed" are reported.

Exploratory Analysis: A follow-up Pearson correlation analysis was conducted to understand the potential role of digital infrastructure in this shift. The analysis revealed that the relationship between internet access and dropout rates changed dramatically between periods.

Pre-Pandemic: There was no significant relationship between internet access and dropout rates (r = 0.057, p = .642).

Post-Pandemic: A significant positive correlation emerged (r = 0.295, p = .015). The results are summarized in Figure A.4.2 in the appendix.

This suggests that internet access transitioned to becoming a contextually important factor, specifically during the crisis. The positive direction of the correlation indicates that in the post-pandemic period, states with higher internet access reported higher dropout rates. This is likely not a direct causal relationship but rather a proxy effect: states with better digital infrastructure may have also implemented more robust reporting systems and digital learning interventions. Their efforts were likely successful at a macro level (contributing to the overall reduction in dropouts), but their improved systems may have also more accurately captured dropout data, and the specific nature of the digital shift may have initially exacerbated access inequalities, making the measured disparity more visible.

Visual Evidence:
The core finding is illustrated in the error bar chart (Figure 3.3a), which shows the mean dropout rate for each period with 95% confidence intervals. The non-overlapping confidence intervals provide clear visual confirmation of the statistically significant difference between the two group means.

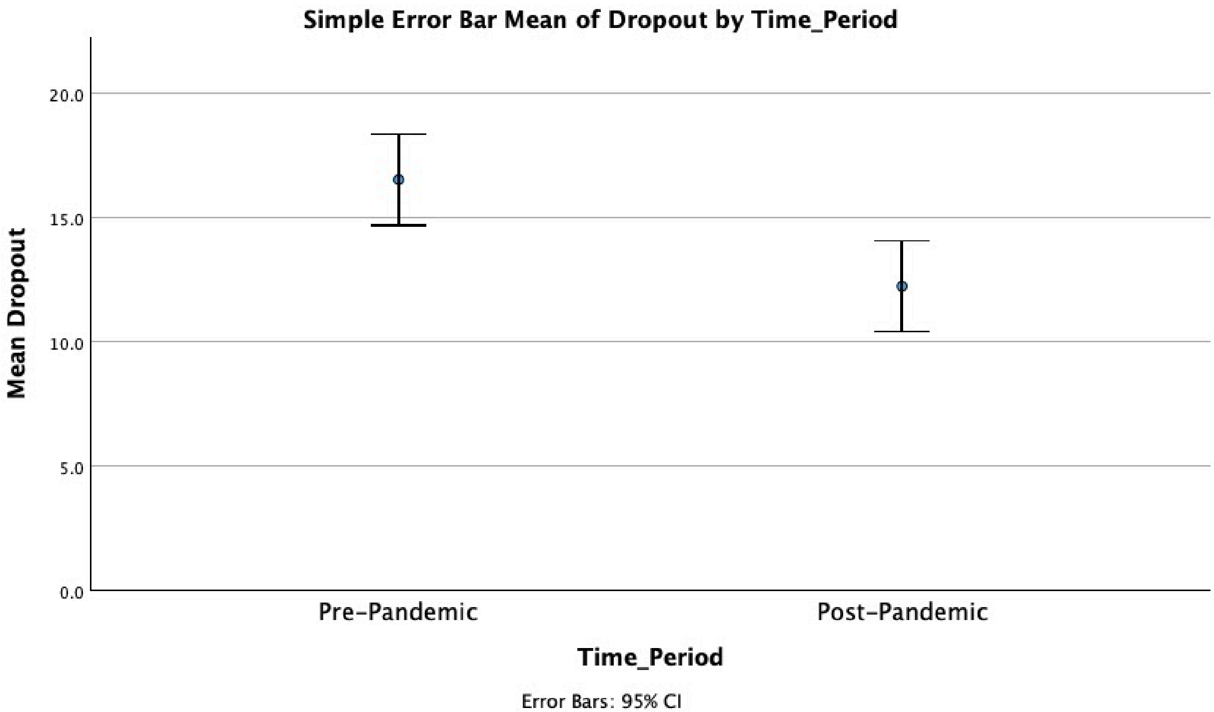


Figure 3.3a: Error Bar Chart of Mean Dropout Rate by Pandemic Period

The results fully support the Pandemic Impact Hypothesis (H_a). We reject the null hypothesis (H_0). The COVID-19 pandemic period was associated with a statistically significant and substantial reduction in student dropout rates across Indian states. While economic well-being remains the primary determinant of dropout rates (as established in H_2), the emergence of a significant correlation with internet access post-pandemic suggests that digital infrastructure played a critical and complex new contextual role during the specific crisis period. This highlights the situation-dependent importance of digital infrastructure, not just as a tool for learning but also as a factor in how educational outcomes are measured and understood during a systemic shock.

4: Conclusion, Recommendations, and Limitations

This study set out to investigate the complex relationship between economic fluctuations and educational outcomes in India during the unprecedented COVID-19 pandemic. Through a multi-tiered analysis, the research yielded three central findings that challenge simplistic narratives and offer nuanced insights for policy and scholarship.

First, at the national policy level, the analysis of the Fiscal Prioritization Hypothesis (H_1) revealed a commendable resilience. Contrary to the expectation that education expenditure would be cut during economic contraction, the government effectively shielded the education budget. Despite a severe GDP decline of -5.78% in 2020, education expenditure as a percentage of GDP remained stable, demonstrating a conscious policy decision to prioritize educational investment during a fiscal crisis.

Second, the test of the Economic Primacy Hypothesis (H_2) confirmed that a state's economic well-being is the most powerful determinant of student retention. The regression analysis demonstrated that higher per capita income is robustly associated with lower dropout rates, a relationship that held even after controlling for digital infrastructure and temporal trends. The finding that internet access was not a statistically significant predictor in this multi-year model underscores that digital solutions, on their own, cannot overcome deep-seated socioeconomic disparities.

Third, and perhaps most surprisingly, the test of the Pandemic Impact Hypothesis (H_3) revealed a significant reduction in dropout rates in the post-pandemic period. This unexpected finding was further illuminated by an exploratory analysis, which showed that the relationship between internet access and dropout rates changed fundamentally. While it was not a significant factor before the pandemic, a significant positive correlation emerged post-pandemic. This suggests the pandemic altered the educational landscape, activating digital infrastructure as a critical contextual factor. The positive correlation likely indicates that states with higher internet access mounted more robust digital learning and reporting initiatives. While these efforts contributed to the overall reduction in dropouts, they may have also measured disparities more accurately or revealed initial inequalities in digital access, highlighting its complex role as a double-edged sword during the crisis.

In conclusion, the pandemic acted not merely as a disruptor but as a catalyst that revealed the layered nature of educational resilience. The findings collectively paint a picture where macroeconomic policy stability and underlying state-level economic strength form the foundational pillars of educational resilience, while digital infrastructure serves as a crucial situational lever and indicator of state response capacity that is activated during acute crises.

4.2 Policy and Strategic Recommendations

Based on these findings, the following recommendations are proposed:

Safeguard Education Expenditure During Downturns: Policymakers should institutionalize the observed practice of protecting education budgets during economic contractions. This research provides empirical evidence that such shielding is not only possible but critical for maintaining educational continuity and equity during crises.

Prioritize Economic Development for Educational Equity: Investments aimed at reducing dropout rates must extend beyond the education sector. Policies focused on broad-based economic development, poverty reduction, and addressing regional income inequality are essential, as economic well-being is the primary driver of educational retention.

Invest in Digital Infrastructure as a Crisis Mitigation Tool: While not a panacea, digital infrastructure proved to be a vital situational asset. Investments in expanding internet access and digital learning platforms should be framed and prioritized as a critical strategy for educational crisis preparedness, ensuring the system is resilient in the face of future shocks.

Adopt a Differentiated Policy Approach: Recognize that the drivers of educational outcomes are not uniform. Policy must distinguish between long-term, foundational investments (economic development) and targeted, situational tools (digital infrastructure), deploying each strategically according to the context.

4.3 Limitations and Avenues for Future Research

While this study provides valuable insights, its limitations point toward productive avenues for future research:

Measurement of Digital Access: A key limitation concerns the measurement of digital infrastructure. India conducts a full population census only every ten years; the 2021 census was postponed. Consequently, it was not possible to calculate precise internet penetration rates (e.g., users per 100 people) for each state-year. The analysis instead used the absolute number of internet subscribers. To mitigate the skew inherent in absolute figures from states with large populations, this variable was log-transformed for the regression analysis. While this transformation normalizes the distribution and improves model fit, it means that the coefficient for internet access reflects the logarithmic change in the absolute number of users rather than a percentage point change in penetration rates, which may limit the intuitive interpretability of its effect size.

Limited National-Level Sample Size: The analysis of national fiscal policy (H_1) was constrained by the short time series (n=4), necessitating a descriptive rather than inferential approach. Future research could employ a cross-national comparative study to achieve greater statistical power for testing this relationship.

Focus on Aggregate Dropout Rates: This study examined overall dropout rates. Future research should disaggregate this data by gender, socio-economic status, and rural/urban location to uncover the differential impact of economic shocks on these subgroups.

Proximate vs. Root Causes: The study establishes a strong correlation between income and dropout rates, but cannot definitively identify the precise mechanistic pathways (e.g., ability to afford devices, need for child labor, parental education). Qualitative follow-up studies are recommended to explore these root causes.

Short-Term Perspective: The study covers the immediate crisis and recovery period (2018-2021). Longitudinal tracking is essential to determine whether the reduction in dropout rates and the emergent role of digital access are sustained trends or temporary phenomena.

Despite these limitations, this study successfully establishes a critical link between economic health and educational outcomes, providing an evidence-based framework for building a more resilient and equitable education system in India and similar economies. Importantly, these limitations contextualize the study's findings. For example, the non-significant result for digital access in H_2 may partly reflect measurement issues with absolute subscriber counts, while the surprising reduction in dropout rates during the pandemic may be influenced by reporting biases or temporary policy measures (e.g., automatic grade promotions). Recognizing these caveats ensures that conclusions are not overstated and provides a realistic foundation for future research

References:

Ayyar, V. (2021). Education and COVID-19 in India: A systematic review of challenges and resilience strategies. *Journal of Education and Development, 5*(2), 45–62.

Government of India, Ministry of Education. (2020). Unified District Information System for Education Plus (UDISE+) report 2019-20. http://udiseplus.gov.in

Government of India, Ministry of Education. (2021). Unified District Information System for Education Plus (UDISE+) report 2020-21. http://udiseplus.gov.in

International Telecommunication Union. (2021). *Measuring digital development: Facts and figures 2021*. https://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx

Reserve Bank of India. (2021). Handbook of statistics on Indian states 2020-2021. https://rbi.org.in/Scripts/AnnualPublications.aspx?head=Handbook%20of%20Statistics%20on%20Indian%20States

Telecom Regulatory Authority of India. (2021). The Indian telecom services performance indicators January - March, 2021. https://www.trai.gov.in/release-publication/reports/performance-indicators-reports

UNESCO. (2020). *Global education monitoring report 2020: Inclusion and education: All means all*. United Nations Educational, Scientific, and Cultural Organization. https://unesdoc.unesco.org/ark:/48223/pf0000373718

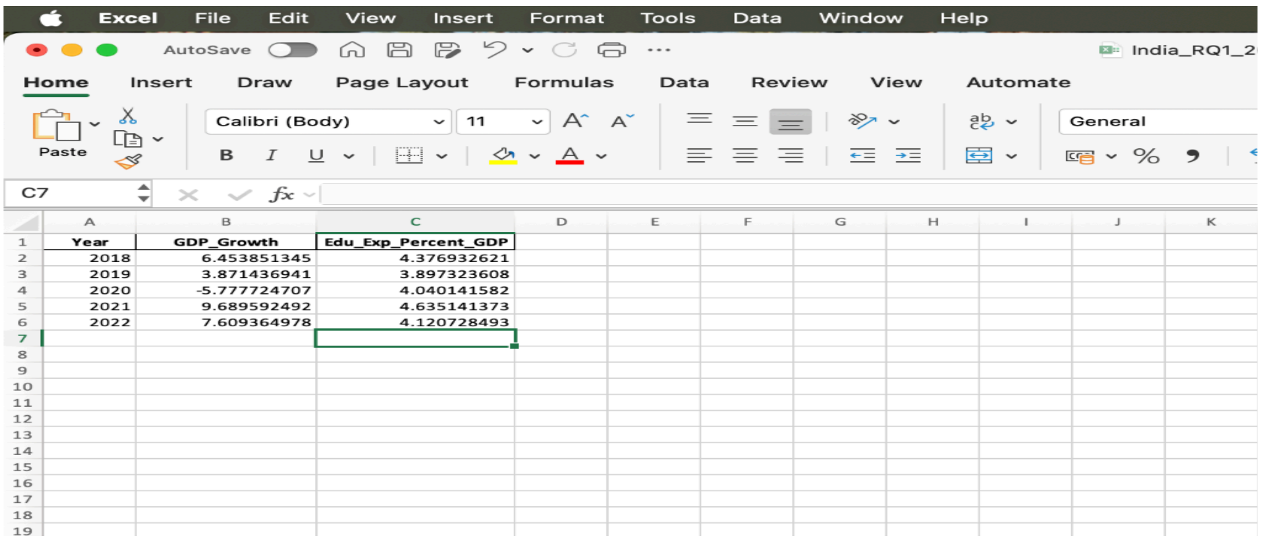
World Bank. (2022). Learning losses from COVID-19 in India: A systematic review. World Bank South Asia Region. https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099125303072236903/P1742520c53c1f06e09ac200dfdc6b3c3c3

Link to our dataset & outputs, raw and processed. https://drive.google.com/file/d/1D4q8srGnYhJR71aIN5459WxSuo_EwOGR/view?usp=sharing

Appendix A: Verification of Data Analysis

A.1: National-Level Data for Time-Series Analysis (H_1)

This is the national-level dataset used for the descriptive time-series analysis of the relationship between GDP growth and education expenditure using Excel.

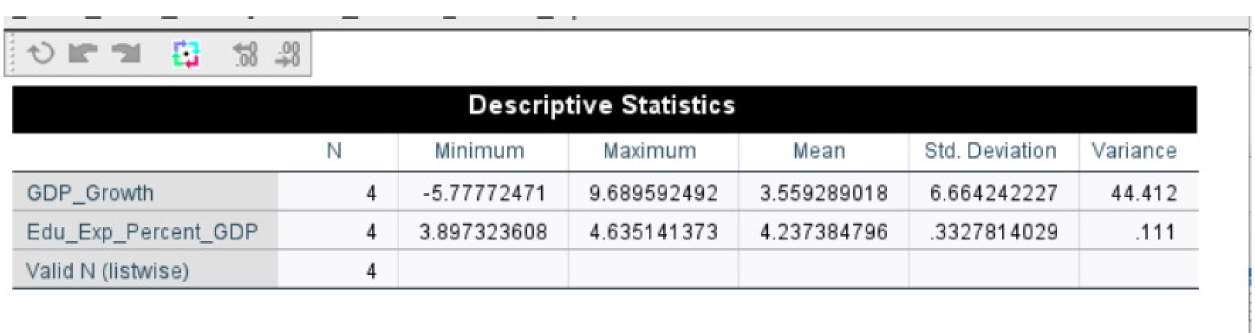


Year	GDP_Growth	Edu_Exp_Percent_GDP
2018	6.453851345	4.376932621
2019	3.871436941	3.897323608
2020	-5.777724707	4.040141582
2021	9.689592492	4.635141373
2022	7.609364978	4.120728493

Figure A.1(a): National Macroeconomic and Education Expenditure Data (2018-2021)

Source: World Bank World Development Indicators. This data was used to generate Figure 3.1 in the main report.

The table below, provides the descriptive statistics for the two national-level variables used in the analysis for H_1. This Excel output quantitatively summarizes the extreme volatility in GDP growth and the stability in education expenditure.



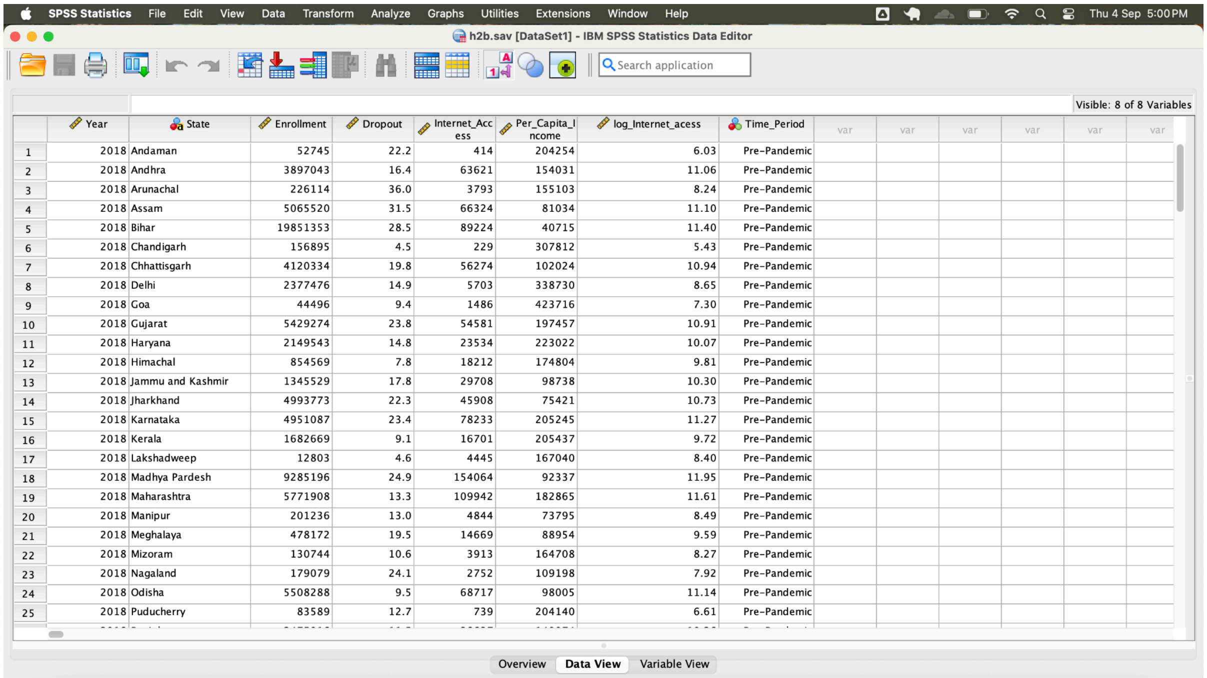
Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
GDP_Growth	4	-5.77772471	9.689592492	3.559289018	6.664242227	44.412
Edu_Exp_Percent_GDP	4	3.897323608	4.635141373	4.237384796	.3327814029	.111
Valid N (listwise)	4					

Figure A.1(b): Excel Output for Descriptive Statistics for National-Level Variables

A.2 State-Level Data Structure and Transformation

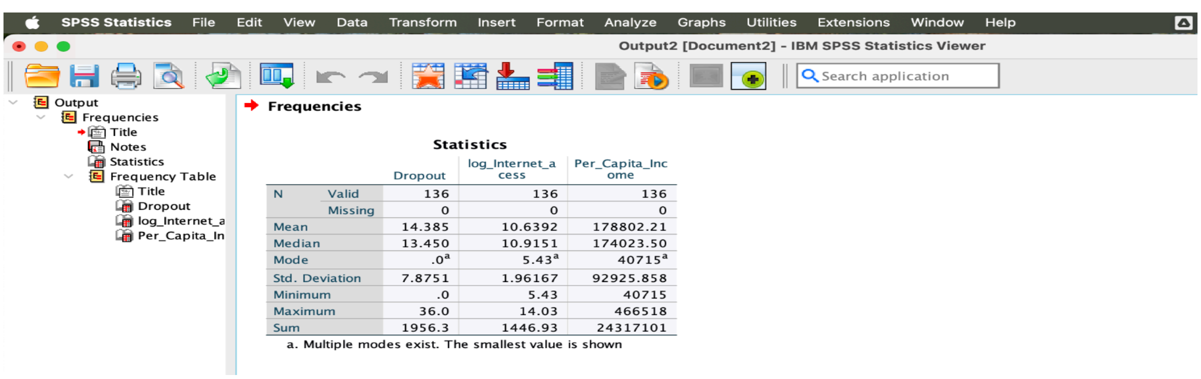
The state-level analysis was conducted on a balanced panel dataset containing N = 136 cases (34 states/union territories across 4 years). An excerpt of the dataset is shown below.



	Year	State	Enrollment	Dropout	Internet_Access	Per_Capita_Income	log_Internet_acess	Time_Period
1	2018	Andaman	52745	22.2	414	204254	6.03	Pre-Pandemic
2	2018	Andhra	3897043	16.4	63621	154031	11.06	Pre-Pandemic
3	2018	Arunachal	226114	36.0	3793	155103	8.24	Pre-Pandemic
4	2018	Assam	5065520	31.5	66324	81034	11.10	Pre-Pandemic
5	2018	Bihar	19851353	28.5	89224	40715	11.40	Pre-Pandemic
6	2018	Chandigarh	156895	4.5	229	307812	5.43	Pre-Pandemic
7	2018	Chhattisgarh	4120334	19.8	56274	102024	10.94	Pre-Pandemic
8	2018	Delhi	2377476	14.9	5703	338730	8.65	Pre-Pandemic
9	2018	Goa	44496	9.4	1486	423716	7.30	Pre-Pandemic
10	2018	Gujarat	5429274	23.8	54581	197457	10.91	Pre-Pandemic
11	2018	Haryana	2149543	14.8	23534	223022	10.07	Pre-Pandemic
12	2018	Himachal	854569	7.8	18212	174804	9.81	Pre-Pandemic
13	2018	Jammu and Kashmir	1345529	17.8	29708	98738	10.30	Pre-Pandemic
14	2018	Jharkhand	4993773	22.3	45908	75421	10.73	Pre-Pandemic
15	2018	Karnataka	4951087	23.4	78233	205245	11.27	Pre-Pandemic
16	2018	Kerala	1682669	9.1	16701	205437	9.72	Pre-Pandemic
17	2018	Lakshadweep	12803	4.6	4445	167040	8.40	Pre-Pandemic
18	2018	Madhya Pardesh	9285196	24.9	154064	92337	11.95	Pre-Pandemic
19	2018	Maharashtra	5771908	13.3	109942	182865	11.61	Pre-Pandemic
20	2018	Manipur	201236	13.0	4844	73795	8.49	Pre-Pandemic
21	2018	Meghalaya	478172	19.5	14669	88954	9.59	Pre-Pandemic
22	2018	Mizoram	130744	10.6	3913	164708	8.27	Pre-Pandemic
23	2018	Nagaland	179079	24.1	2752	109198	7.92	Pre-Pandemic
24	2018	Odisha	5508288	9.5	68717	98005	11.14	Pre-Pandemic
25	2018	Puducherry	83589	12.7	739	204140	6.61	Pre-Pandemic

Figure A.2(a): SPSS Excerpt of the Cleaned Dataset

Source: UDISE Ministry of Education, RBI, and TRAI. This data was used for all state-level analyses. The variable Internet_Access was log-transformed to correct for severe positive skewness.



Frequencies

Statistics

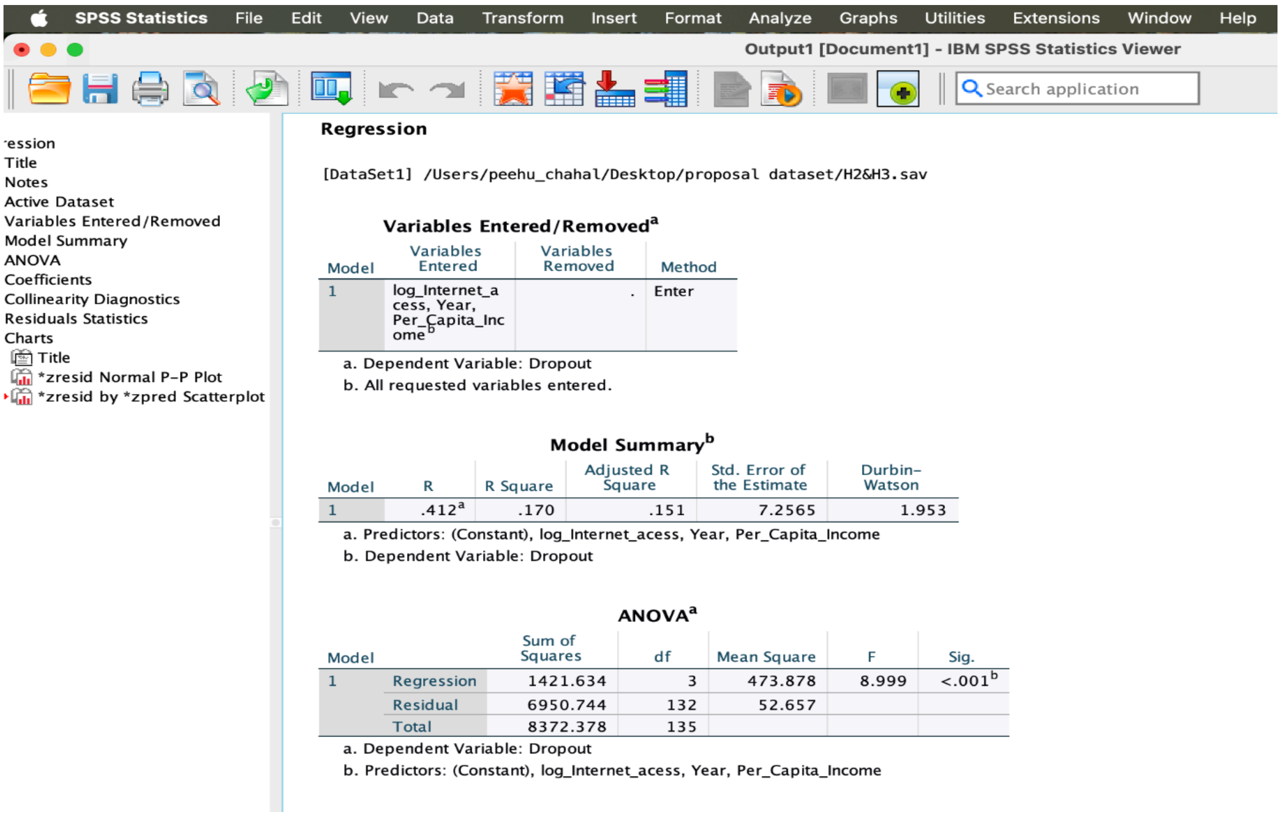
		Dropout	log_Internet_acess	Per_Capita_Income
N	Valid	136	136	136
	Missing	0	0	0
Mean		14.385	10.6392	178802.21
Median		13.450	10.9151	174023.50
Mode		.0[a]	5.43[a]	40715[a]
Std. Deviation		7.8751	1.96167	92925.858
Minimum		.0	5.43	40715
Maximum		36.0	14.03	466518
Sum		1956.3	1446.93	24317101

a. Multiple modes exist. The smallest value is shown

Figure A.2(b): Cleaned Dataset Overview with Statistics

A.3.1: SPSS Output for Multiple Linear Regression Assumptions(H_2)

This is the statistical output for the multiple regression Assumptions model testing the Economic Primacy Hypothesis.



Regression

[DataSet1] /Users/peehu_chahal/Desktop/proposal dataset/H2&H3.sav

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	log_Internet_acess, Year, Per_Capita_Income[b]	.	Enter

a. Dependent Variable: Dropout
b. All requested variables entered.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.412[a]	.170	.151	7.2565	1.953

a. Predictors: (Constant), log_Internet_acess, Year, Per_Capita_Income
b. Dependent Variable: Dropout

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1421.634	3	473.878	8.999	<.001[b]
	Residual	6950.744	132	52.657		
	Total	8372.378	135			

a. Dependent Variable: Dropout
b. Predictors: (Constant), log_Internet_acess, Year, Per_Capita_Income

Figure A.3.1(a) SPSS Output for Multiple Linear Regression Assumptions(H_2)

Continued below..

ession
Title
Notes
Active Dataset
Variables Entered/Removed
Model Summary
ANOVA
Coefficients
Collinearity Diagnostics
Residuals Statistics
Charts
Title
*zresid Normal P-P Plot
*zresid by *zpred Scatterplot

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	Upper Bound	Collinearity Statistics Tolerance	VIF
1	(Constant)	4004.489	1132.564		3.536	<.001	1764.165	6244.814		
	Per_Capita_Income	-2.097E-5	.000	-.247	-2.843	.005	.000	.000	.830	1.204
	Year	-1.975	.561	-.281	-3.521	<.001	-3.085	-.865	.984	1.016
	log_Internet_acess	.247	.348	.062	.711	.479	-.441	.935	.839	1.192

a. Dependent Variable: Dropout

Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions (Constant)	Per_Capita_Income	Year	log_Internet_acess
1	1	3.795	1.000	.00	.01	.00	.00
	2	.189	4.482	.00	.69	.00	.03
	3	.016	15.292	.00	.28	.00	.97
	4	1.509E-7	5015.277	1.00	.01	1.00	.00

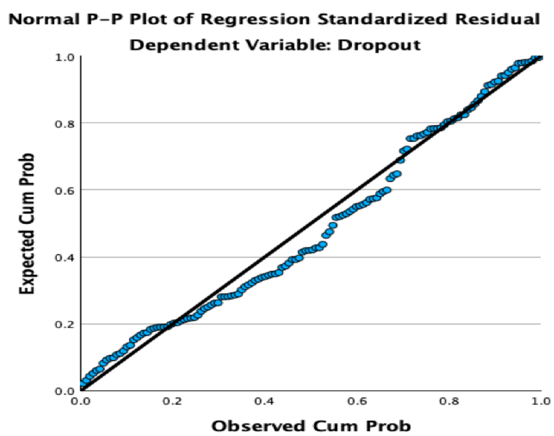
a. Dependent Variable: Dropout

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	4.524	20.430	14.385	3.2451	136
Residual	-14.6948	19.0946	.0000	7.1754	136
Std. Predicted Value	-3.039	1.863	.000	1.000	136
Std. Residual	-2.025	2.631	.000	.989	136

a. Dependent Variable: Dropout

Charts



rplot

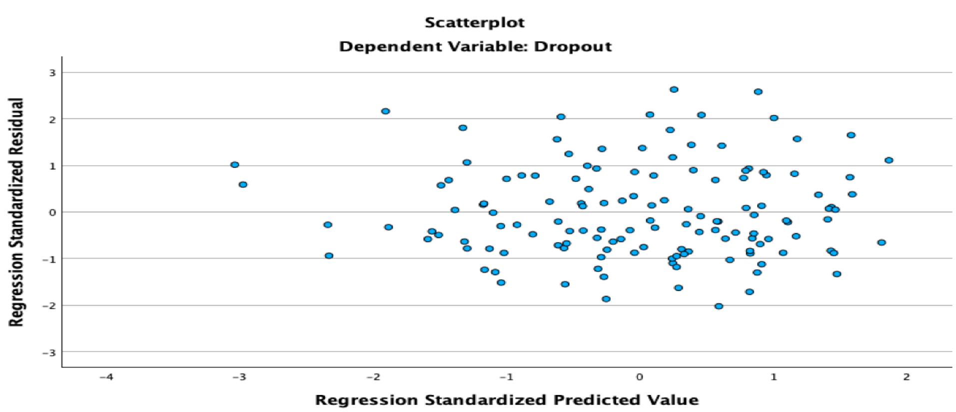
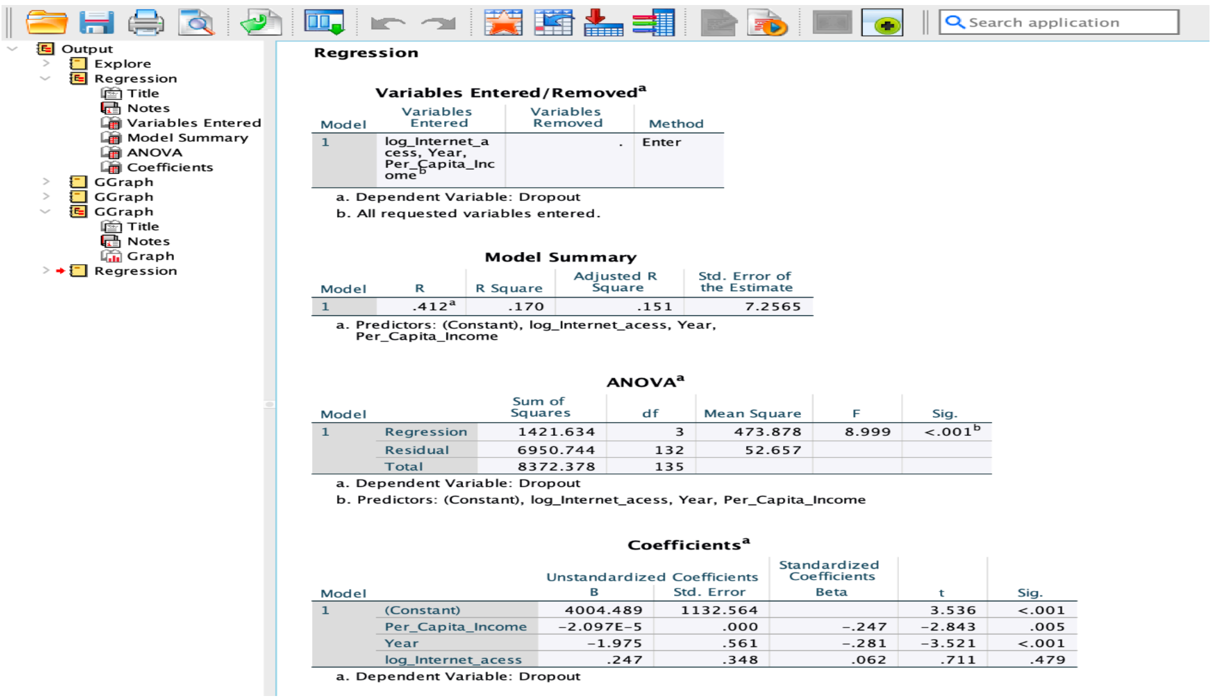


Figure A.3.1(b) SPSS Output for Multiple Linear Regression Assumptions(H_2)

A.3.2: SPSS Output for Multiple Linear Regression (H_2)

This is the statistical output for the multiple regression model testing the Economic Primacy Hypothesis.



Regression

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	log_Internet_acess, Year, Per_Capita_Income[b]	.	Enter

a. Dependent Variable: Dropout
b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.412[a]	.170	.151	7.2565

a. Predictors: (Constant), log_Internet_acess, Year, Per_Capita_Income

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1421.634	3	473.878	8.999	<.001[b]
	Residual	6950.744	132	52.657		
	Total	8372.378	135			

a. Dependent Variable: Dropout
b. Predictors: (Constant), log_Internet_acess, Year, Per_Capita_Income

Coefficients[a]

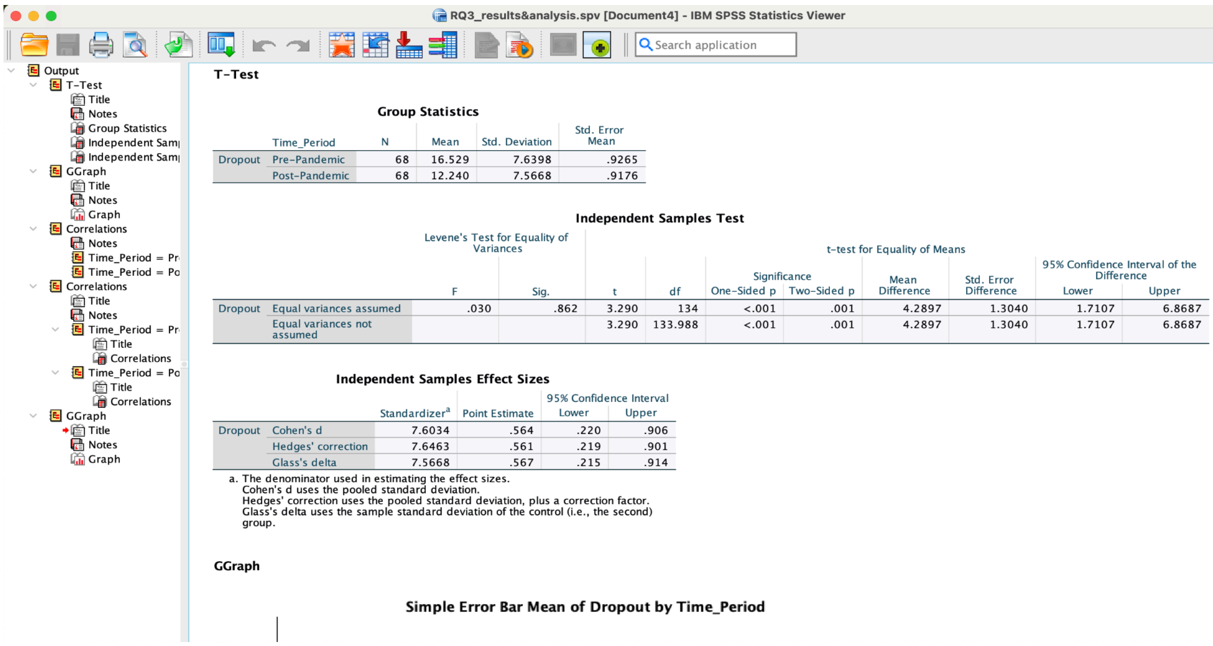
Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	4004.489	1132.564		3.536	<.001
	Per_Capita_Income	-2.097E-5	.000	-.247	-2.843	.005
	Year	-1.975	.561	-.281	-3.521	<.001
	log_Internet_acess	.247	.348	.062	.711	.479

a. Dependent Variable: Dropout

Figure A.3.2: SPSS Output for Multiple Linear Regression (H_2) Coefficients

A.4: SPSS Output for Independent Samples T-Test and Correlation (H_3)

This is the statistical output from IBM SPSS Statistics for the t-test and correlation comparing pre-pandemic and post-pandemic dropout rates.



RQ3_results&analysis.spv [Document4] - IBM SPSS Statistics Viewer

T-Test

Group Statistics

	Time_Period	N	Mean	Std. Deviation	Std. Error Mean
Dropout	Pre-Pandemic	68	16.529	7.6398	.9265
	Post-Pandemic	68	12.240	7.5668	.9176

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance				95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
Dropout	Equal variances assumed	.030	.862	3.290	134	<.001	.001	4.2897	1.3040	1.7107	6.8687
	Equal variances not assumed			3.290	133.988	<.001	.001	4.2897	1.3040	1.7107	6.8687

Independent Samples Effect Sizes

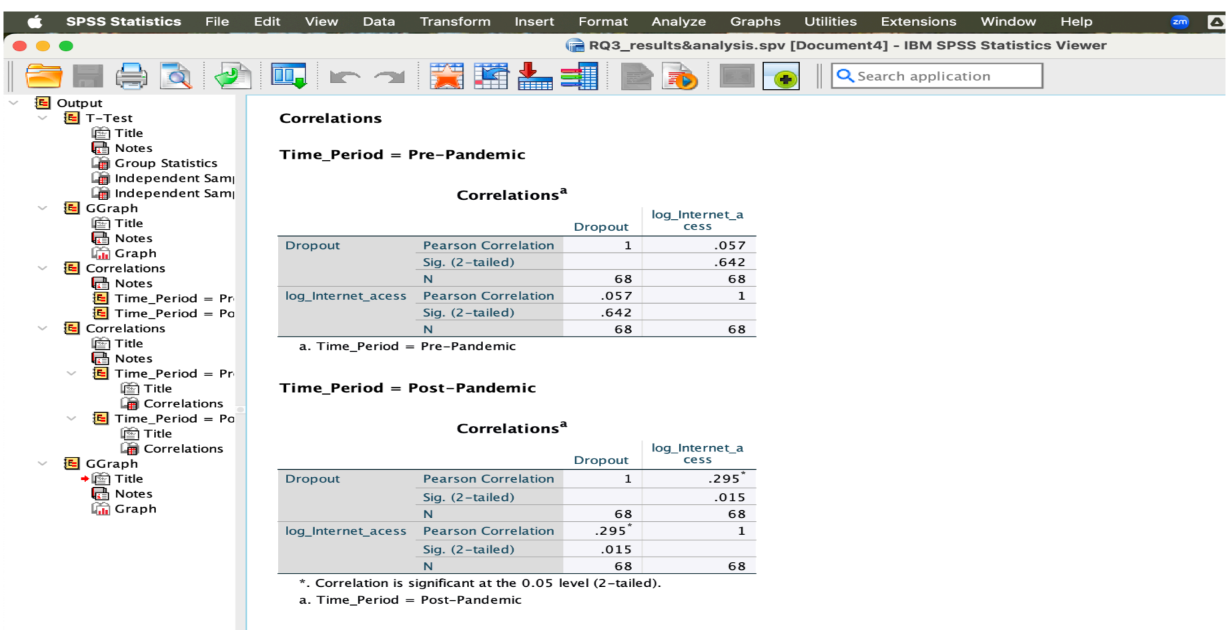
		Standardizer[a]	Point Estimate	95% Confidence Interval Lower	95% Confidence Interval Upper
Dropout	Cohen's d	7.6034	.564	.220	.906
	Hedges' correction	7.6463	.561	.219	.901
	Glass's delta	7.5668	.567	.215	.914

a. The denominator used in estimating the effect sizes.
Cohen's d uses the pooled standard deviation.
Hedges' correction uses the pooled standard deviation, plus a correction factor.
Glass's delta uses the sample standard deviation of the control (i.e., the second) group.

GGraph

Simple Error Bar Mean of Dropout by Time_Period

Figure A.4(a): SPSS Output for Independent Samples T-Test



RQ3_results&analysis.spv [Document4] - IBM SPSS Statistics Viewer

Correlations

Time_Period = Pre-Pandemic

Correlations[a]

		Dropout	log_Internet_acess
Dropout	Pearson Correlation	1	.057
	Sig. (2-tailed)		.642
	N	68	68
log_Internet_acess	Pearson Correlation	.057	1
	Sig. (2-tailed)	.642	
	N	68	68

a. Time_Period = Pre-Pandemic

Time_Period = Post-Pandemic

Correlations[a]

		Dropout	log_Internet_acess
Dropout	Pearson Correlation	1	.295*
	Sig. (2-tailed)		.015
	N	68	68
log_Internet_acess	Pearson Correlation	.295*	1
	Sig. (2-tailed)	.015	
	N	68	68

*. Correlation is significant at the 0.05 level (2-tailed).
a. Time_Period = Post-Pandemic

Figure A.4(b): SPSS Output for Bivariate Correlations(Pre & Post Pandemic Period: 2018-20 & 2021-22)

Appendix B: Variable Construction Note

B.1: Measurement of Digital Infrastructure

A significant limitation concerns the measurement of digital infrastructure. Due to the postponement of the 2021 Indian census, it was not possible to calculate precise internet penetration rates (e.g., subscribers per 100 people) for each state and year. Consequently, the absolute number of internet subscribers was used as the best available proxy.

To mitigate the skew inherent in absolute figures from states with vastly different populations, this variable was log-transformed (using the natural logarithm, ln) for the regression analysis to normalize its distribution. The resulting variable, log_Internet_Access, was used in all multivariate models. This means the coefficient reflects the relationship between the logarithmic change in the absolute number of subscribers and dropout rates, rather than a percentage point change in penetration rates. Future research should seek to replicate these findings with per-capita penetration rates once the data becomes available.