PANEL DATA REGRESSION ANALYSIS IN JMP: EDUCATION SPENDING AND LITERACY ACROSS INDIAN STATES (2005–2022)

Client Profile:

The client was a public policy postgraduate student preparing a dissertation on the effectiveness of government education expenditure across Indian states. Their objective was to determine whether annual education spending had a statistically significant effect on literacy rate improvements from 2005 to 2022.

Client's Problem:

The client collected panel data for 20 Indian states over 18 years. However, they faced challenges in:

1. Structuring data appropriately for panel analysis in JMP.
2. Identifying whether fixed or random effects were appropriate.
3. Understanding interaction effects between rural population percentage and education spending.
4. Writing a technically sound report in Harvard format with clear policy implications.

My Role:

- Econometric Analyst (Panel Data in JMP)
- Visualization & Report Developer (Harvard format)
- Interpretation Coach (for dissertation defense)

Tools Used:

- JMP Pro (Panel Regression with Mixed Model platform)
- Excel (initial data cleaning)
- Zotero (for referencing)

Step 1 – Data Structuring and Import into JMP

- Transformed raw Excel files into a stacked panel format with identifiers for State_ID and Year.
- Verified balanced panel (no missing years) and created dummy variables for regional controls (e.g., North vs South India).
- Created interaction term: Education_Spending * Rural_Pop_Percent.

Step 2 – Descriptive Analysis and Visualization

- Created **panel-specific box plots** for education spending and literacy rate trends across states.
- Used **JMP's Trellis Plots** to display trends by region.
- Performed **pairwise correlation analysis** and scatter matrix for early visual inspection of associations.

Step 3 – Model Specification and Execution

- Ran initial OLS regression ignoring state clustering to generate baseline coefficients.
- Used JMP's **Mixed Model platform** to run panel regressions with:
    - **Fixed Effects (State & Year)**
    - **Random Effects (State)**
    - **Interaction Term**
- Compared models using **AIC and BIC** and found that the **Fixed Effects Model** had the best performance and interpretability.

Step 4 – Diagnostics and Robustness Checks

- Tested for:
    - **Heteroskedasticity** using White's test (robust SEs applied)
    - **Autocorrelation** using residual plots and Durbin-Watson (no issues found)
    - **Multicollinearity** via VIF (all < 4)
- Ran model both with and without interaction term to test its significance.

Step 5 – Interpretation of Results

- Key finding: Education spending had a **positive and statistically significant impact** on literacy rate ($p = 0.01$).
- Interaction term with rural population share was also significant ($p = 0.03$), indicating that **returns to education spending were higher in rural-heavy states**.
- Adjusted $R^2 = 0.71$.

Final Report Delivery

Delivered a **12-page Harvard-style report**, including:

- Research Background & Policy Context
- Panel Data Structure and Model Choice
- JMP Outputs: Coefficient Tables, Fixed Effects Plots
- Visualizations: Time Trends, Regional Comparisons
- Interpretation with Policy Relevance
- Limitations and Suggestions for Further Research

The report was structured to meet academic submission standards and included clear policy suggestions based on the econometric results.

Key Metrics and Project Summary

Item	**Value**
Time Span	2005–2022
No. of Units	20 Indian states
Panel Type	Balanced Panel
Model Used	Fixed Effects Panel Regression
Interaction Term	Spending × Rural Population (%)
Significant Predictors	Education Spending, Interaction Term
Adjusted R^2	0.71
Output	Harvard-style report + JMP graphs

Completion Time	10 days