

ANALYZING ACADEMIC PERFORMANCE TRENDS IN NATIONAL TEST SCORES USING MATLAB

1. Objective

The project aimed to detect long-term patterns and anomalies in student performance data using national assessment scores. The focus was on subject-level trends over time across multiple regions to assist educational planning and policy revisions.

2. Client and Use Case

Client Type:

Public education board overseeing K-12 education in an English-speaking country.

Use Case:

The board needed a structured analysis of assessment results from 2014 to 2023 to evaluate the impact of curriculum changes and identify regions with persistent underperformance.

3. Dataset Overview

Source:

National standardized assessment results (Grades 3, 5, and 8)

Time Span:

2014–2023

Sample Size:

- ~1.2 million student records
- ~3,000 schools
- 5 major subjects: Math, Reading, Science, Social Studies, Writing

Key Fields:

- Student ID (anonymized)
- School ID and Region
- Year
- Subject
- Score (0–100 scale)

- Demographics: gender, language spoken at home, socioeconomic index

4. Methodology

4.1 Data Preparation

- Merged CSV files across 10 years using readtable and innerjoin
- Cleaned and standardized score formats
- Aggregated data at the region-year-subject level for trend analysis

4.2 Statistical Techniques

1. Time Series Decomposition:

- Used tsmovavg and detrend to separate trend, seasonal, and residual components
- Applied smoothing (Exponential Moving Average) to highlight gradual performance shifts

2. ANOVA and Regression:

- Conducted Two-Way ANOVA to check interaction effects of region and year on average scores
- Built linear models using fitlm to quantify the slope of score changes over time

3. Outlier Detection:

- Identified regional anomalies in specific years using Z-score thresholds
- Used boxplots per subject-year for visual confirmation

5. Key Findings

- **Math and Science:** steady national improvement (~0.5 points/year)
- **Writing:** declining performance in 3 out of 5 regions
- **Reading gaps** widened between low-SES and high-SES regions post-2019
- 2021–2022 scores showed pandemic-related dips, with slow recovery in 2023

Statistical Significance:

- ANOVA: Region*Year p-value < 0.001
- Linear regression slope (Reading, Region C): -1.18 points/year

6. Visualizations

- **Trend plots** (subject-wise, region-wise) with confidence intervals
- **Heatmaps** showing subject-level performance changes by year
- **Boxplots** comparing SES-based score distributions across years
- **Bar charts** of top and bottom performing regions

All visualizations were generated using MATLAB's plot, boxplot, heatmap, and bar functions.

7. Deliverables

- Region-wise trend analysis report (PDF, 38 pages)
- Subject-wise dashboards built using MATLAB App Designer
- PowerPoint briefing deck for board presentation
- Cleaned and labeled .mat dataset for internal use
- Executive summary with policy-focused recommendations

8. Strategic Impact

- Curriculum teams initiated review of writing instruction techniques in underperforming regions
- Funding proposals drafted for Reading interventions in low-SES areas
- Public dashboards were proposed for transparency in academic progress reporting
- Pandemic learning loss reports were included in education ministry submissions

9. Tools and Features Used

- MATLAB R2023a
- Statistics and Machine Learning Toolbox
- App Designer for region-specific data exploration
- Time Series functions for smoothing and forecasting
- Exported graphics via exportgraphics for publication use