EVALUATING THE RELATIONSHIP BETWEEN ONLINE LEARNING HOURS AND EXAM PERFORMANCE: A MULTIPLE LINEAR REGRESSION STUDY

Objective:

The primary goal of this project was to assess whether students' academic performance (final exam scores) could be predicted using their weekly online learning hours, previous GPA, and participation in discussion forums. The regression results would support the client's case for scaling their online learning platform in higher education institutions.

Client Requirements:

- Use synthetic or anonymized real-world educational data
- Include three predictors: weekly learning hours, GPA, and discussion participation
- Address issues of outliers and multicollinearity
- Generate an APA-style report with clear conclusions
- Provide recommendations for online education design

Data Source and Structure:

A dataset of 600 student records was collected from an online university's internal LMS. It included the following variables:

- Final exam score (continuous)
- Weekly hours spent on platform (continuous)
- Cumulative GPA (continuous)
- Number of discussion posts per week (continuous)
- Gender and course type (for exploratory analysis only)

Methodology:

1. **Preprocessing and Cleaning:**

- Checked for and removed cases with incomplete GPA or exam score data
- Standardized weekly learning hours and discussion posts
- Outliers were assessed via standardized residuals (±3 SD), with 7 removed

2. **Regression Model Specification:**

$$\text{ExamScore}_i = \beta_0 + \beta_1 \cdot \text{LearningHours}_i + \beta_2 \cdot \text{GPA}_i + \beta_3 \cdot \text{Posts}_i + \epsilon_i$$

3. **Analysis Tool:**
   - Regression analysis conducted using **SPSS 28**
   - Outputs formatted for APA reporting
   - Plots generated using **Excel** and **R (ggplot2)** for clarity
4. **Model Validation:**
   - VIF values were all below 2.1, ruling out multicollinearity
   - Durbin-Watson = 1.92, indicating no autocorrelation
   - Homoscedasticity and normality of residuals confirmed visually and statistically

Key Results:

- **Weekly Learning Hours** was a significant predictor (**$\beta = 2.14$, $p < 0.001$**): every additional hour improved exam score by approx. 2.14 points
- **Cumulative GPA** had the strongest impact (**$\beta = 5.87$, $p < 0.001$**)
- **Forum Posts** had a modest but significant effect (**$\beta = 1.02$, $p = 0.04$**)
- **Adjusted $R^2 = 0.62$**, indicating the model explains 62% of the variance in final scores

Deliverables:

- **14-page regression report** including tables, interpretations, and APA-style citations
- **Regression diagnostics** included as an appendix
- **Visual dashboards** with coefficient impact charts and student segment distributions
- **Summary handout** for non-technical decision-makers

Client Outcome and Feedback:

The client, an edtech startup, used the insights to prioritize features in their product roadmap. They expanded emphasis on forum engagement tools and modular content time tracking, citing our report in their internal pitch to investors. They praised the clarity of the regression visuals and the professional layout of the final deliverables.