

EMPLOYEE PERFORMANCE EVALUATION USING CORRELATION AND REGRESSION IN EXCEL

1. Background and Problem Statement:

An IT services company sought to evaluate the **impact of training hours on employee performance** to justify its L&D (Learning and Development) investment. While HR had maintained detailed records of training sessions and performance ratings for the past year, no statistical analysis had been conducted. The goal was to quantify the relationship between training hours and appraisal scores and to build a simple **performance prediction model** using Excel, accessible to HR executives.

2. Objectives:

- Determine whether a correlation exists between training hours and performance scores
- Use linear regression to model and predict employee performance based on training
- Create an interactive Excel report to explore department-wise performance data
- Provide recommendations for targeted training investments

3. Methodology:

3.1 Dataset Overview

- **Sample Size:** 200 employees (5 departments)
- **Variables Included:**
 - Employee ID
 - Department
 - Total training hours (last 12 months)
 - Final performance rating (scale of 1 to 5)
 - Years of experience

3.2 Data Preparation

- Handled missing training data by imputing with department-wise median
- Normalized training hours using z-scores for visual comparisons

- Used **named ranges** and **structured tables** for dynamic referencing in formulas

3.3 Statistical Analysis

- **Pearson correlation coefficient (CORREL)** to measure relationship strength
- **Simple linear regression (LINEST and TREND)** for performance prediction
- Visualized data with scatter plots and trendlines
- Created **interactive department filters using slicers**

4. Results and Interpretation:

4.1 Correlation Analysis

- Pearson correlation (r) between training hours and performance score: **0.62**
- Indicates a **moderate to strong positive relationship**
- Highest $r = 0.72$ in QA department; lowest $r = 0.49$ in Support

4.2 Regression Model

- **Linear regression equation:** Performance Score = $2.31 + 0.018 \times \text{Training Hours}$
- **$R^2 = 0.39$:** 39% of the variability in performance score is explained by training hours
- Slope = 0.018 means every additional 10 hours of training corresponds to a 0.18 point increase in rating (on a 5-point scale)

4.3 Excel Visualization

- Created scatter plots with regression lines by department
- Built dropdown filters and conditional formatting to highlight outliers
- Added a **forecast calculator**: user enters training hours → predicted rating auto-generates

5. Recommendations:

- Set minimum 30 training hours annually for departments with lower correlation
- Prioritize high-impact modules for teams with poor regression fit (e.g., Support)
- Use regression model to identify **underperformers relative to training hours** for coaching
- Track R^2 and slope yearly to evaluate training ROI

6. Deliverables in Excel:

- Clean, interactive dashboard with:
 - Performance vs. training scatter plots
 - Summary stats table by department
 - Employee-level prediction calculator
- Separate tab with full regression breakdown and formula walkthrough

7. Stakeholder Relevance:

Academic:

- Practical example of applying correlation and regression using native Excel tools
- Suitable for teaching statistical modeling and visualization in Excel-based analytics courses

Corporate:

- Equips HR and L&D teams with a **low-code framework** for linking training to performance
- Enables simple modeling, scenario testing, and performance forecasting
- Provides a foundation for broader workforce analytics adoption