

PREDICTING EMPLOYEE PERFORMANCE FROM TRAINING HOURS AND ENGAGEMENT LEVELS USING SPSS MULTIPLE LINEAR REGRESSION

1. Background and Problem Statement:

An IT services firm implemented multiple upskilling programs to improve team productivity and innovation. However, despite increased investment in training, leadership observed inconsistent improvement in employee performance ratings. The HR analytics team was tasked with understanding whether variables like training hours, engagement level, and years of experience were statistically related to performance. A regression-based modeling approach was chosen to test these relationships in SPSS.

2. Objectives:

- To determine if training hours and engagement levels predict employee performance
- To control for years of experience and department differences
- To check for interaction effects between training and engagement
- To create a model that HR can use for future workforce planning

3. Methodology:

3.1 Data Collection:

- **Sample Size:** 320 employees from 4 departments (Tech, HR, Sales, Support)
- **Data Collected:**
 - Dependent Variable: Annual Performance Rating (0–100 scale)
 - Independent Variables:
 - Training Hours (numeric)
 - Engagement Score (scale: 1–10)
 - Years of Experience (numeric)
 - Department (dummy-coded: HR, Sales, Support; Tech = reference)
 - Interaction Term: Training Hours × Engagement Score

3.2 SPSS Procedure:

- Descriptive statistics for all variables
- Dummy coding for categorical department variable
- Regression Model 1: Main effects only
- Regression Model 2: Adds interaction term
- Diagnostic checks for:
 - Multicollinearity using VIF
 - Linearity and residual plots
 - Outlier analysis using standardized residuals and Cook's distance

4. Results:

Model Fit (Model 2):

- $R^2 = 0.71$, Adjusted $R^2 = 0.70$
- $F(6, 313) = 129.52$, $p < 0.001$

Significant Predictors:

- Training Hours: $\beta = 0.32$, $p < 0.001$
- Engagement Score: $\beta = 0.43$, $p < 0.001$
- Years of Experience: $\beta = 0.21$, $p = 0.004$
- Interaction Term: $\beta = 0.15$, $p = 0.03$

Department Effects:

- HR: $\beta = -2.8$, $p = 0.09$
- Sales: $\beta = -5.1$, $p = 0.01$
- Support: $\beta = -1.2$, $p = 0.40$

Diagnostics:

- VIF values < 3 for all variables
- Residuals showed no major violations of homoscedasticity or normality
- No extreme outliers detected

5. Interpretation and Insights:

- Engagement score is the strongest predictor of performance, even more than training hours
- Training is effective, but its impact increases when employees are already engaged
- Employees with more experience tend to perform better, although marginally
- Sales department employees have significantly lower performance ratings after controlling for other variables
- Interaction analysis indicates that engagement boosts the effectiveness of training—less engaged employees do not benefit as much from longer training durations

6. Recommendations:

- Enhance pre-training engagement through gamification or recognition programs
- Target more interactive, role-specific training in Sales department
- Use regression model to identify low-performing individuals with low training hours and engagement scores for proactive intervention
- Consider quarterly engagement assessments to track real-time readiness for training

7. Future Work:

- Integrate 360-degree feedback scores into performance prediction
- Explore non-linear models to capture diminishing returns of high training hours
- Use panel data over three years to track changes within employees

8. Stakeholder Relevance:

Academic:

- Suitable example for HR analytics coursework and interaction effects in regression
- Demonstrates use of dummy variables, interaction terms, and diagnostics in SPSS

Corporate:

- Helps HR teams build data-informed learning and development strategies
- Practical tool for managers to allocate training budgets based on predicted impact