Application of Statistical Modeling for the Personalization and Optimization of Training in a Collegiate Lacrosse Team

Zaín de Jesús Vázquez-Ramírez[1], Armando Olarte-Vargas[2]

Meritorious Autonomous University of Puebla

Introduction.

The growing availability of sports data underscores the need for tools that transform raw information into actionable insights for performance optimization. In lacrosse, expected goals (xG), compared with actual goals, provide a robust framework to evaluate training effectiveness [1]. This study analyzes 60 post-tournament sessions, modeling the relationship between xG and shot outcomes using polynomial and logistic regressions. Time-series analysis tracked performance trends, while clustering techniques grouped sessions based on Goals/xG efficiency, revealing recurring behavioral patterns [2]. This integrated, data-driven approach offers a structured methodology to guide strategic training adjustments and enhance both individual and team performance.

Methodology.

Data were collected from 60 training sessions of a collegiate lacrosse team, conducted three times per week during the post-tournament period. Expected goals (xG) metrics were recorded, and heatmaps were generated to analyze both the efficiency and spatial distribution of plays on the field [1]. The relationship between xG and shots was modeled using a second-order polynomial regression, complemented by logistic regression to capture nonlinear behaviors. Additionally, clustering was applied to the time-series data to group sessions according to performance patterns, enabling the identification of recurring configurations and the evaluation of trends in offensive efficiency.

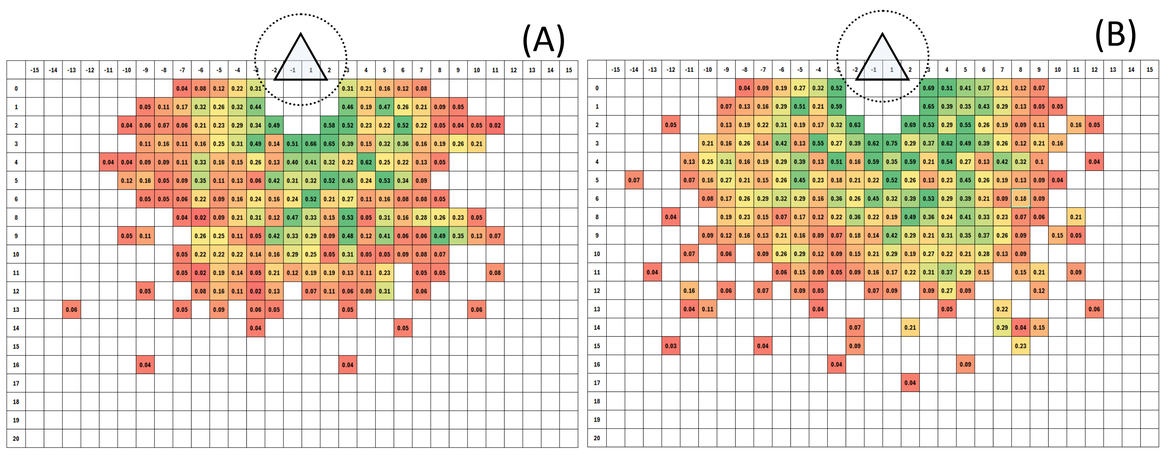

Figure 1. Heatmap of xG distribution in the scoring zone: (A) start of the study, (B) final week.

Results.

The goal-per-shot ratio relative to xG was modeled using a second-order polynomial regression estimated via ordinary least squares, complemented by a logistic component fitted through maximum likelihood. The resulting fit revealed a nonlinear relationship: conversion rates were low at low xG values, reached a maximum at intermediate ranges, and decreased at high xG values, consistent with a potential statistical saturation effect.

Time-series analysis of 60 post-tournament sessions tracked goals/xG evolution [2]. From the second month, k-means clustering on normalized series (validated with the silhouette index) grouped sessions with similar behaviors, enabling a structured interpretation of offensive dynamics.

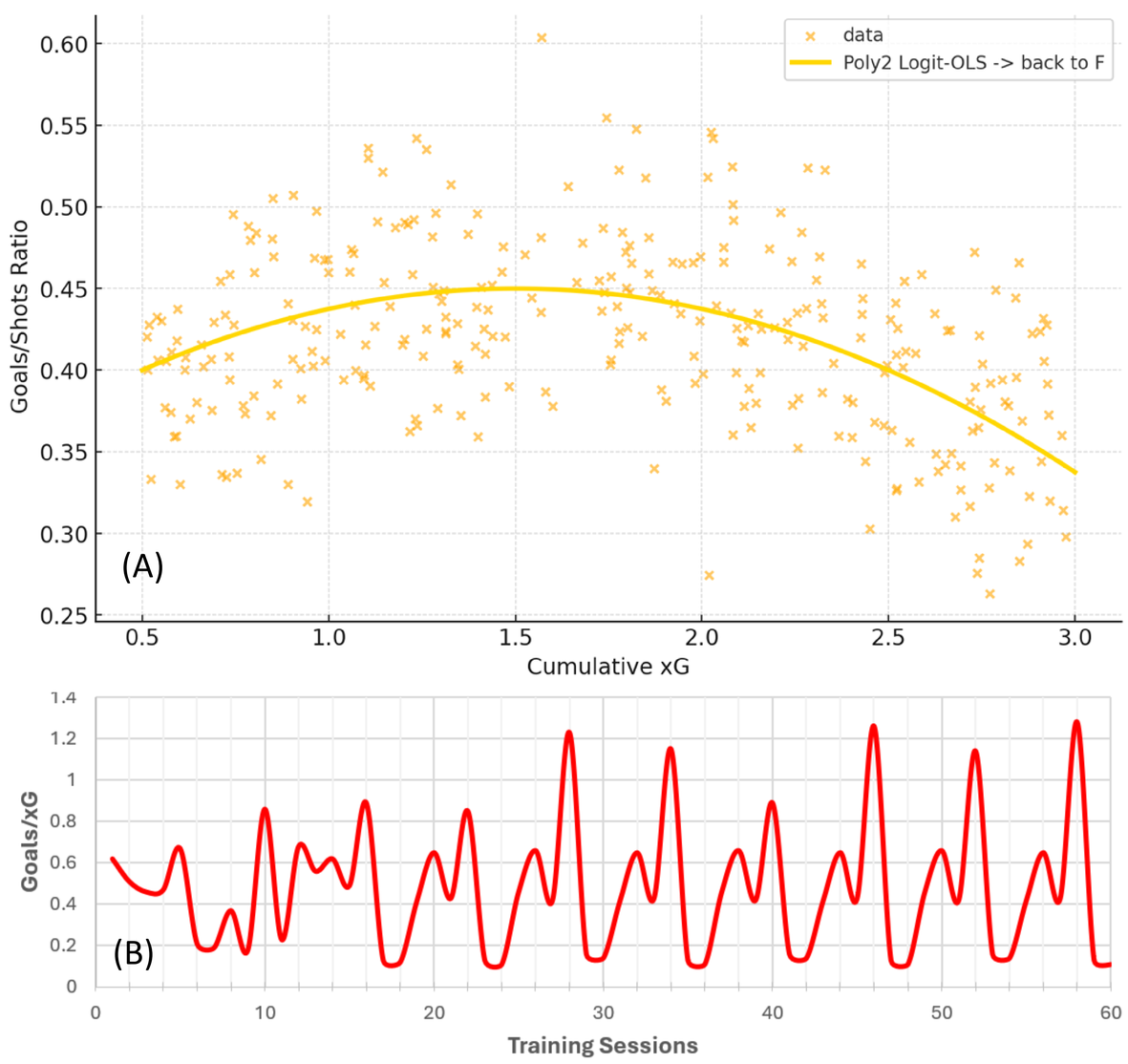

Figure 2. Results visualizations: (A) Polynomial regression curve, (B) Time-series of goals/xG evolution.

Conclusions.

In this study, the goals/xG relationship exhibited nonlinear behavior, peaking at intermediate ranges and decreasing at extreme values, consistent with a statistical saturation effect. Time-series analysis of 60 post-tournament sessions enabled the characterization of offensive performance evolution. The integration of second-order polynomial regression, logistic regression, and clustering techniques facilitated the identification of recurring patterns and the grouping of sessions with homogeneous behaviors. This structured, data-driven approach provides a robust methodological framework for optimizing training planning and enhancing individual and team performance.

References.

[1] Myers, B. R., Burns, M., Coughlin, B. Q., & Bolte, E. (2021). On the development and application of an expected goals model for lacrosse. *The Sport Journal*, 24.

[2] Komitova, R., Raabe, D., Rein, R., & Memmert, D. (2023). Time series data mining for sport data: A review. International Journal of Computer Science in Sport, 21(2), 17–31.