Overview

In this study, we conducted a comparative analysis of players' performance in both international and club soccer games to assess whether their respective performances exhibit similarities. The motivation for this investigation stems from the common criticism players face from the general public for allegedly falling short of expectations, particularly in international competitions.

Our project focuses on scrutinizing and comparing player performance across both domains – club and international games. To achieve this, we employed the Poisson distribution as a statistical tool, known for its application in discrete probability modeling. By leveraging the Poisson distribution, our objective extends beyond a mere performance comparison; we aim to use it as a predictive tool to forecast players' performances in our upcoming project.

Dataset

The dataset for this project comprises two CSV files. The first file encompasses club data, detailing player performances in club games, while the second file contains data from the same years, capturing player performances in international games. These datasets were sourced from Wikipedia's official website, www.wikipedia.org.

The CSV file dedicated to international games features twelve players listed in rows, with 8 columns each presenting data on their country, appearances, and goals over three years (2006, 2010, and 2014). Similarly, the second CSV file, focusing on club data, also has the same twelve players in rows, with 8 columns reflecting their respective clubs, appearances, and goals during the same three-year span.

Analysis

In general, a player's performance is evaluated primarily by how many goals he/she scored. Hence, we have chosen the same approach to measure their performance. Their performance comparison between both (international and club) games by looking only at their goals will not be an ideal approach since their appearances between those games might be different. Hence, during analysis, we have calculated the goal per appearance to have more accuracy in our conclusion. Goal per appearance is calculated by dividing total goals scores of an individual player by the total number of games they appeared in. In other words, it is an average goal scored per their appearance i.e., lambda (λ). Lambda has been calculated by using below formula:

$$\lambda = \frac{Total\ Goals}{Total\ no.\ of\ Appearances}$$

Under Poisson distribution of statistics, Lambda is the mean (average) value as well as the variance. Hence, goal per appearances i.e., Lambda (λ) of each individual player for both club and international games has been calculated during data analysis. In order to compare the calculated lambda values between two games, a 'modified z-scores' has been calculated using below formula:

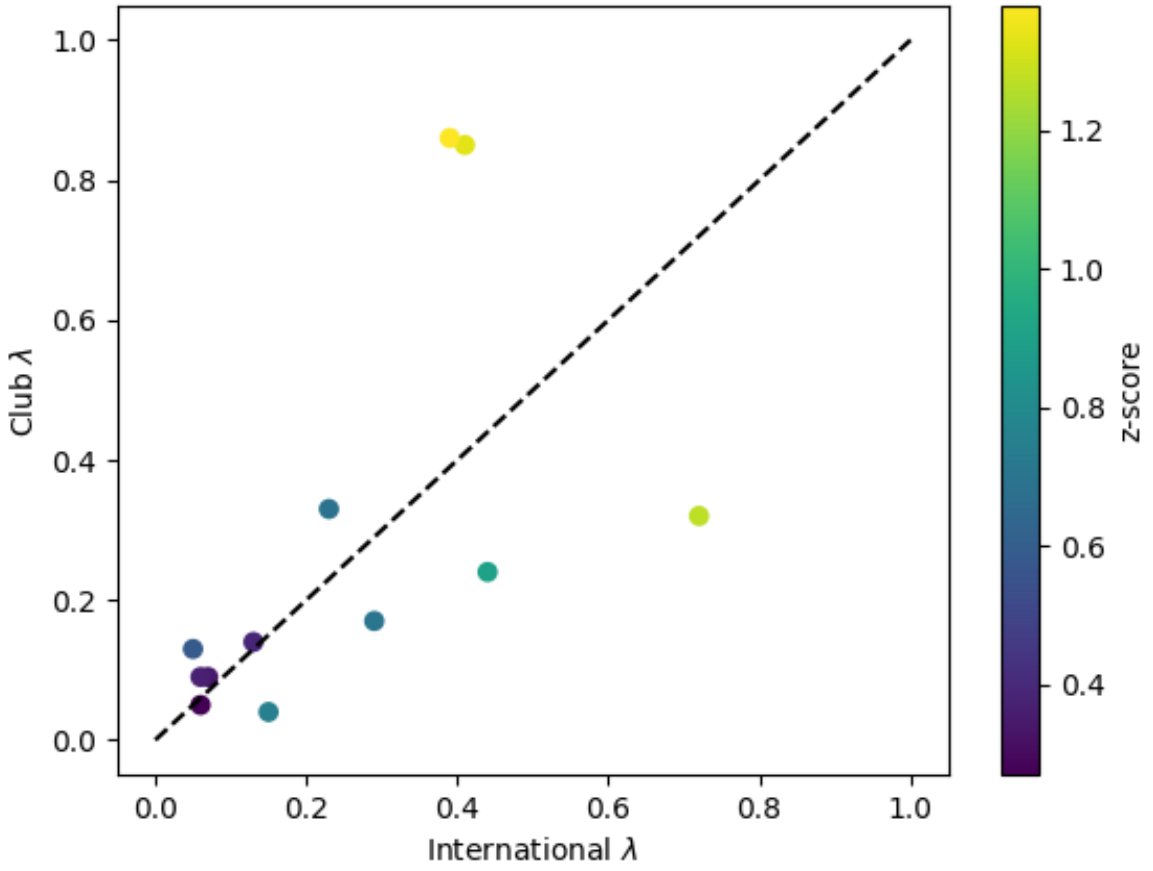
$$Modified\ z - score = \frac{Max\left(Lambda_{club}, Lambda_{Intl}\right)}{Min\left(Lambda_{club}, Lambda_{Intl}\right)}$$

As we utilize the modified z-score as our evaluation metric, a z-score exceeding 1 will be interpreted as indicating a significant difference in performance, while a z-score less than or equal to 1 will be regarded as not significant.

When executing the Python code, the following output is generated:

Player	Lambda (Club)	Lambda (Intl)	Z-Score	Remarks
Cristiano Ronaldo	0.85	0.41	1.33	>1
Lionel Messi	0.86	0.39	1.38	>1
Sergio Ramos	0.13	0.05	0.58	<1
Rafael Márquez	0.04	0.15	0.75	<1
Xavi	0.09	0.07	0.34	<1
Andrés Iniesta	0.14	0.13	0.39	<1
Andrés Guardado	0.09	0.06	0.37	<1
Wesley Sneijder	0.33	0.23	0.69	<1
Tim Cahill	0.24	0.44	0.90	<1
Miroslav Klose	0.32	0.72	1.27	>1
Philipp Lahm	0.05	0.06	0.27	<1
Bastian Schweinsteiger	0.17	0.29	0.70	<1

Examining the above z-scores, we can infer that only three players out of twelve (Cristiano Ronaldo, Lionel Messi, and Miroslav Klose) exhibit significantly different performances, as indicated by their respective z-scores (1.33, 1.38, and 1.27, all exceeding 1).



The accompanying scatter plot, generated using Python, displays Lambda_Intl on the x-axis and Lambda_Club on the y-axis, with dot colors denoting the z-scores. Notably, only three outliers are evident in the scatter plot.

Conclusion

Given that nine out of twelve players have z-scores below 1, we conclude that there is no significant difference between the players' performances in club and international games. This analysis holds significant relevance for various stakeholders. Countries participating in the World Cup can leverage insights into their players' club performances to strategically manage their teams. Likewise, soccer clubs can assess a player's performance in the World Cup, to make decisions on recruitment or team management. For individuals engaged in sports betting, this analysis provides a basis for making more informed predictions. Sports commentators stand to benefit by scrutinizing players' performances in both club and international games, allowing them to identify noteworthy players for commentary during matches.

Looking ahead, with the possibility of collecting more comprehensive data from reliable sources, researchers could conduct more in-depth studies, considering factors such as a player's assists, shots on target, and more. Furthermore, researchers could delve into specific roles, such as goalkeepers and defenders, exploring their performances based on metrics such as saves made and fouls committed.