Olympic medal prediction combining deterministic and stochastic factors

Fang Guo[1], Xiyun Chen[2], Jiayue Wang[3]

[1]Communication University of China, Beijing, 100024, China;

[2]North China Electric Power University, Beijing, 102206, China;

[3]university of Glasgow, Glasgow, G12 8QQ, United Kingdom;

Fang Guo, Xiyun Chen, Jiayue Wang are co-first authors

Abstract. This paper analyzes several influencing factors by constructing a predictive model for the Olympic medal table, providing valuable insights to assist national Olympic committees in optimizing their Olympic strategies. In the preprocessing stage, the study analyzed each country's performance over the past four Olympic Games regarding gold and total medals. Countries were grouped into two categories—developed and developing sports nations — and visualized through Principal Component Analysis (PCA). This classification is a foundational element for the predictive model in the first question. The study converted Olympic medals from past Games into corresponding scores in the first question. Calculating Pearson correlation coefficients showed that the various Olympic sports can broadly be categorized into ball and non-ball. A linear regression model was then constructed to predict medals in ball and non-ball sports using the performance data from the past three Olympic Games as independent variables. In the second question, we analyzed the relationships between coaching assignments and the performance of the women's volleyball and gymnastics teams for three pairs of countries: China-USA and Romania-USA. Through analysis, we concluded that the "coach effect" exists, with national team scores strongly correlated with outstanding coaches. Additionally, the contribution of exceptional coaches to a nation's total medal count exceeded 0.5. In the third question, we pointed out that the prediction of Olympic medals requires careful consideration of the leader effect, host effect, candidate effect, excellent coach effect, etc., providing a reference for the Olympic Committee to designate competition rules.

Keywords: k-means; elbow method; Pearson Correlation Coefficients; Host Effects; Probabilistic.

1. Introduction

The Olympic Games, held every four years, are set to return in 2028 in Los Angeles, USA. As a global platform for countries and athletes to showcase their athletic prowess and compete for medals across diverse events, the Games have inspired extensive research into accurately predicting medal counts for each nation in the past. However, this task is becoming increasingly complex due to various uncertainties, as it needs to consider personal, social, event-specific, and time-specific factors that influence outcomes. To address this challenge, in this paper, a composite predictive model will be developed to estimate medal counts based on deterministic factors.

The Olympic Games, a prestigious quadrennial international sports event, has become a key arena for global competition. Medal rankings reflect national strength. This paper constructs a predictive model for Olympic medals, providing valuable insights for national Olympic committees to optimize strategies. Using k-means clustering and the elbow method, combined with principal component analysis (PCA), countries were classified into two groups: sports powerhouses and developing nations, based on their gold and total medal performance in the past four Olympic Games. This clustering forms the foundation for the predictive model.

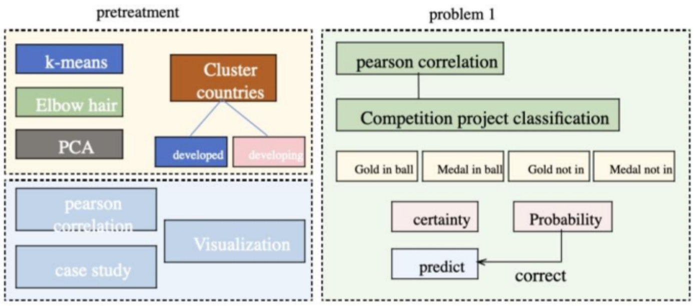


Fig. 1 Model Overview

Olympic medals were converted into scores to analyze the impact of different sports. Pearson correlation coefficients revealed two significant categories: ball and non-ball sports. A linear regression model was constructed using the data from the past three Olympic Games, with additional factors like the "host effect," "leader effect," and "candidate effect" incorporated as probabilistic parameters. Optimized through a simulated annealing algorithm, the model achieved over 90% accuracy in interval predictions.

2. Preparation

2.1 Data Preprocessing

2.1.1 K-Means Clustering: Objective, Algorithm, and Optimal Cluster Selection

The K-Means algorithm partitions n data points into k clusters to minimize the within-cluster variance (also known as the sum of squared errors, SSE). Given a set of observations $(\mathcal{X}1, \mathcal{X}2, \ldots, \mathcal{X}n)$, where each observation is a *d*-dimensional real vector, k-means clustering aims to partition the n observations into l (≤n) sets $S = \{S_1, S_2, \ldots, S_l\}$ to minimize the within-cluster sum of squares (WCSS) (i.e., variance). Formally, the objective is to find:

$$arg_s min \sum_{i=1}^{l} \sum_{X\in S_I} \parallel X - \mu_I \parallel^2 = arg_s min \sum_{i=1}^{l} |s_i| Var S_i \quad (1)$$

Where μ_i is the mean (also called centroid) of points in S_i, i.e.

$$\mu_i = \frac{1}{|S_i|} \sum_{x,y\in S_i} \parallel \mathcal{X} - \mathcal{Y} \parallel^2 \quad (2)$$

The equivalence can be deduced from identity $|S_i| \sum_{X\in S_i} \parallel \mathcal{X} - \mu_i \parallel^2 = \frac{1}{2} \sum_{X,y\in S_i} \parallel \mathcal{X} - \mathcal{Y} \parallel^2$.

Since the total variance is constant, this is equivalent to maximizing the sum of squared deviations between points in different clusters (between-cluster sum of squares, BCSS)[1]. This deterministic relationship is also related to the law of total variance in probability theory[i].

2.1.2 Determining Optimal Clusters with the Elbow Method

The Elbow Method is widely used to determine the optimal number of clusters k. It involves plotting the SSE against different values of k and identifying the point where the rate of decrease slows significantly, forming an "elbow" in the graph. This point represents the optimal k, balancing simplicity and the ability to capture meaningful groupings in the data.

Step 1: Calculate SSE for Different Values of L. For a range of L values (e.g., from 1 to 10), apply K-means clustering to the dataset and calculate the Sum of Squared Errors (SSE) for each L.

$$SSE = \sum_{i=1}^{n} \sum_{j=1}^{l} w^{(i,j)} \|\mathcal{X}i - \mu^j\|_2^2 \quad (3)$$

$\mathcal{X}i$: The i-th data point.

$w^{(i,j)}$: Binary indicator variable (1 if $x^{(i)}$ belongs to the jth cluster, otherwise 0)

μ^j: The center point (centroid) of the j-th cluster.

$\|\mathcal{X}i - \mu^j\|_2^2$: The square of the Euclidean distance between $\mathcal{X}i$ and μ^j.

Step 2: plot the SSE values as a graph, with values on the x-axis and the corresponding SSE values on the y-axis.

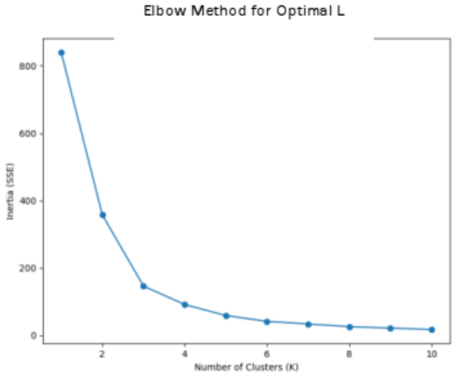


Fig. 2 Elbow method for K-means visualization

Step 3: In the SSE graph, the objective of the number of $O_{\text{Number of Clusters(L)}}$ is to identify where the rate of decrease in SSE slows down. In other words, increasing the number of clusters before this point reduces the SSE, but after this point, the reduction in SSE becomes much less pronounced. This point resembles the "elbow" in a human arm, representing the "bend" in the SSE curve.

Step 4: The L value corresponding to the elbow point is considered the optimal number of clusters for the dataset. In the example above, the Elbow Method shows significant fluctuations in the SSE graph when L=2 and L=3.

2.1.3 PCA and Implementation

PCA (Principal Component Analysis) is a widely used dimensionality reduction algorithm to address this issue. It reduces the number of variables while retaining as much information as possible by transforming correlated variables into a smaller set of uncorrelated ones, called principal components. That enables more efficient and comprehensive analysis of high-dimensional data. This study will do PCA using Eigenvalue Decomposition (Golub-Kahan method).

Step 1: Input dataset $X = \{X_1, X_2, X_3, \ldots, X_n\}$, reduce to t-dimensions.

Step 2: Subtract the mean of each feature (mean-centering).

Step 3: compute the covariance matrix $\frac{1}{n}XX^T$.

Step 4: Perform eigenvalue decomposition of the covariance matrix $\frac{1}{n}XX^T$ to obtain its eigenvalues and eigenvectors.

Step 5: Sort the eigenvalues in descending order and select the top t eigenvalues. Use the corresponding t eigenvectors to form the eigenvector matrix P.

Step 6: Transform the data into the new space spanned by these t eigenvectors: Y = PX.

Algorithm: Golub-Kahan method (EVD)
Parameter: Number of epoch S, Number of dimension t
Input: Data matrix $X \in R^{E\times D}$
Output: Eigenvectors $W \in R^{D\times t}$
$C = XX^T/D$
$U_1TU_1^T = \text{Householder}(C)$
$U_2\Lambda U_2^T = \text{Diagonalization}(T)$
$U = U_1U_2$
$W = X^TU$

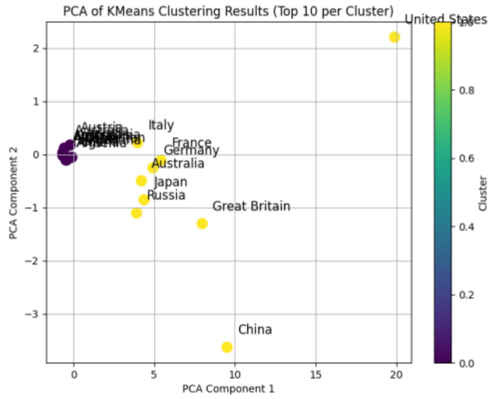


Fig. 3 PCA of K-Means results

This visualization represents the clustering results following PCA dimensionality reduction. The plot's yellow points denote sports powerhouses, while the purple points represent developing sports nations. Notably, countries such as the United States, China, Australia, and Germany are classified as sports powerhouses, reflecting the objectivity and validity of the clustering methodology applied in this study.

Randomly picking three from the list of countries, we can tell from their K-means output results that Burkina Faso's minimal medal counts across all categories position it as a developing sports nation. In contrast, Great Britain's dominant performance, with significantly higher total and gold medals, confirms its status as a sports powerhouse. With moderate medal counts, Cuba performs better than developing nations but lacks the broad dominance of a powerhouse. These distinctions validate the effectiveness and objectivity of categorizing countries into these groups for further analysis and prediction.

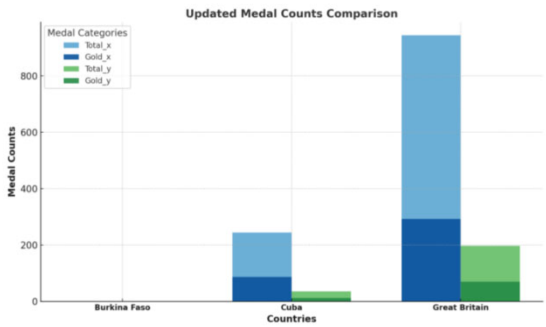


Fig. 4 Medal counts comparison

3. Predict the Number of Medals for Each Country

3.1 Technical Background

3.1.1 Inter-Sport Correlations and Impact on Predictions: Pearson's coefficient

Pearson's correlation coefficient, often denoted by the Greek letter ρ when applied to a population, is referred to as the population correlation coefficient or the population Pearson correlation coefficient[ii]. For a pair of random variables (M, N) (e.g., Height and Weight), the formula for ρ is given as:

$$\rho_{M,N} = \frac{cov(M,N)}{\sigma_M \sigma_N} \tag{4}$$

This formula can also be written in the form of:

$$\rho_{M,N} = \frac{\mathbb{E}[(M-\mu_M)(N-\mu_N)]}{\sigma_M \sigma_N} \tag{5}$$

Based on the fact that:

$$COV(M,N) = \mathbb{E}[(M-\mu_M)(N-\mu_N)] \tag{6}$$

As we analyzed the medal scores across all countries and Games, calculating the correlation coefficients can identify whether performance in one sport is positively or negatively associated with another, providing insights into patterns and dependencies between sports.

The heat map above illustrates Pearson's correlation coefficients between different sports, highlighting the strength and direction of relationships. Red indicates positive correlations, and blue indicates negative correlations. We found the sports with the highest correlations are as follows:

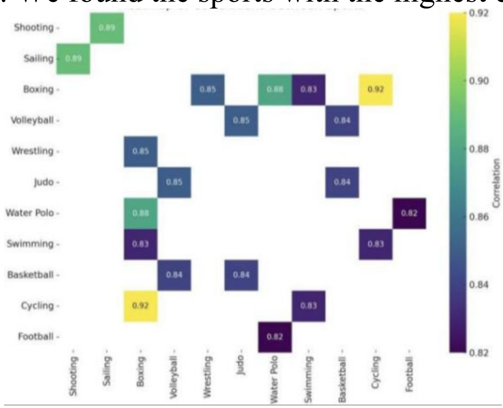


Fig. 5 Pearson's correlation coefficients heat map 2

We observed that, firstly, some sports are significantly correlated with each other. These highly correlated sports can generally be categorized into those requiring high physical exertion, such as cycling and boxing, or ball sports, such as volleyball and basketball. The coefficient analysis reveals that the relationship between different sports and a country's medal counts varies.

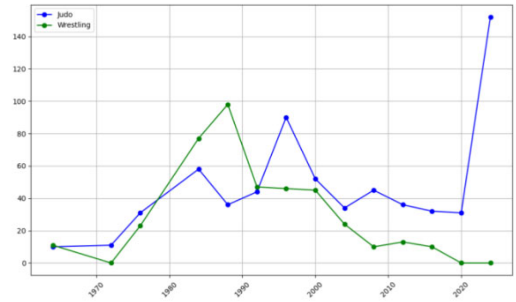


Fig. 6 South Korea's historical Judo and Wrestling scores

The scoring trends show that, apart from 2024, the two sports exhibit similar patterns. Their similar trends over time indicate that performance in these two sports is closely related, reflecting shared factors influencing their success. Also, the scoring trends for volleyball and basketball in the United States across the Olympic Games show high similarity, as shown below.

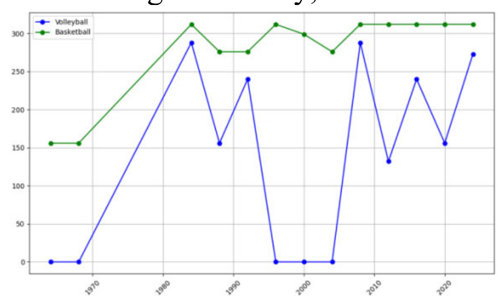


Fig. 7 U.S. Basketball and Volleyball scores

However, as shown in the chart below, judo and basketball in the United States have low degrees of similarity in their trends.

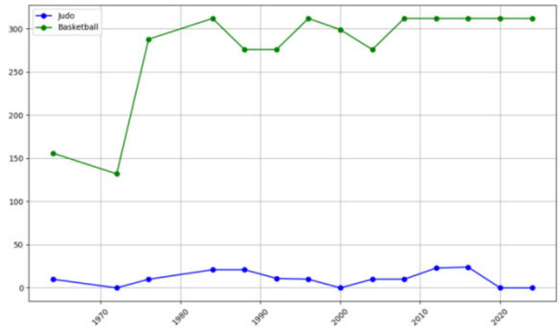


Fig. 8 U.S. Judo and Basketball scores

Based on this analysis, we can propose a hypothesis categorizing Olympic sports into ball and non-ball sports. Within each category, representative factors, such as the mutual relationships between basketball and volleyball or judo and wrestling, play a role.

3.1.2 Host Effect

The host effect typically refers to a phenomenon in sports competitions where the host country achieves exceptional results by leveraging various advantages of being the host[iii]. We calculate the total scores for gold, silver, and bronze medals achieved by different host countries in their respective Olympic Games. Additionally, we compute the total scores for all events and countries in the same games and determine the proportion of the host country's score relative to the total scores of all participating countries, as shown below.

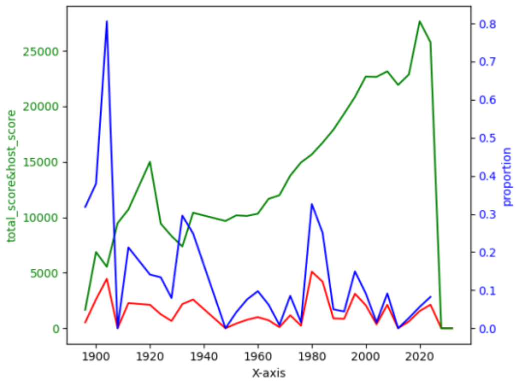


Fig. 9 Host scores and total scores comparison

In the early editions of the Olympic Games, the host effect was highly pronounced, with the proportion reaching as high as 80%. However, as Olympic events evolved, competitions became increasingly fair, and events became more scientifically organized, gradually reducing the host effect. We applied the moving average method to calculate the sliding average of the host country's score proportion using the following formula:

$$proportion_i = \frac{\sum_{r=i}^{F+i-1}(Z_r)}{F} \tag{7}$$

Z_i: The host country's scoring ratio of the nth Olympic Games

F: The size of the sliding window (i.e., the number of terms involved in the calculation)

i: The index of the current term

r: sum index, indicating the index variable used to traverse data when calculating moving average.

With a window size of 4 Olympic Games, the moving average method reduces the impact of random fluctuations on the data. The host effect can generally oscillate downward by observing the proportions. Therefore, we can use time series decomposition methods to predict the future trend of the host effect:

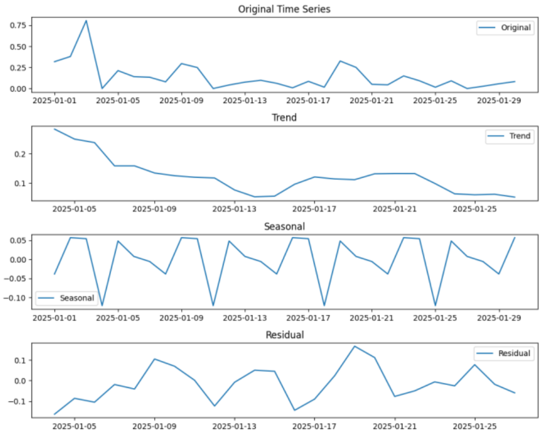


Fig. 10 Time Series Decomposition

The results of the time series decomposition are shown above. Using linear regression to predict the trend component and adding the cyclical component to the prediction results in a rolling manner, we get the expected host effect values for the next four Olympic Games separately: [0.09972629, 0.09335355, 0.0869808, 0.08060806]. Therefore, we adopt 0.09 as the host effect value for the 2028 Olympic Games.

3.1.3 Leadership Effect

When certain athletes from a country participate in multiple Olympic Games, the "leadership effect" may emerge. These athletes, referred to as "leader athletes," represent a nation's core strength. A higher proportion of such leader athletes relative to the total number of athletes from that country indicate more excellent expertise and consistency in a specific field.

Next, we calculate the proportion of leader athletes from different nations to estimate the effect of leaders in other countries.

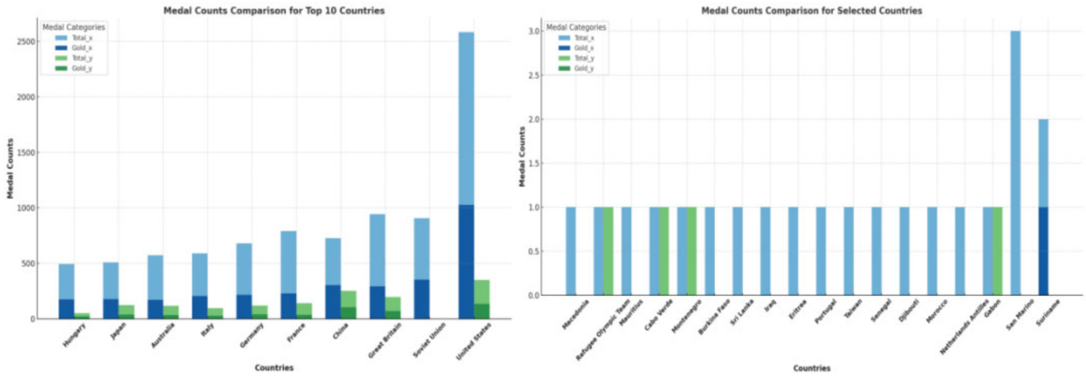


Fig. 11 Medal counts for top and bottom 10 countries

Through comparative analysis, the "leadership effect" is more prevalent in developing sports nations. That is likely because developed countries, in terms of sports performances, generally have more robust sports education systems and a broader talent pool, enabling faster turnover of athletes. That highlights a structural difference in athlete development and retention strategies between developed and developing nations.

3.1.4 Reserve Talent Effect & Talent Increment Effect

We quantify these effects using the number of participants in different events across countries in the same Olympic year and by analyzing changes in participant numbers between consecutive Olympic Games. The proportion of events experiencing a decline is higher after 2000 than before 2000. With reforms introduced by the International Olympic Committee, some events were removed from the Gamesiv, leading to a streamlined athlete pool and decreased Talent Increment Effect.

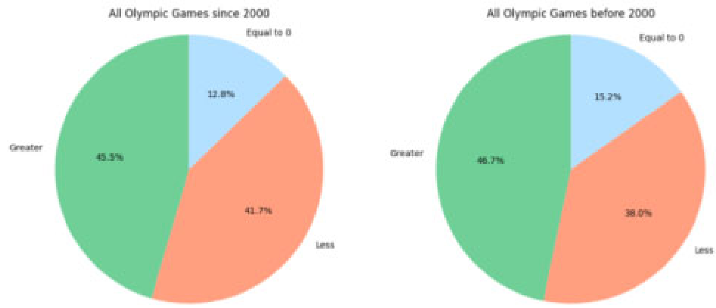


Fig. 12 Changes in participant numbers between consecutive Olympic Games

3.2 Programs and Justification

3.2.1 Regression

The predictive model for answering Problem 1 consists of two components: a deterministic part and a stochastic part:

$$y_{pred} = \hat{y} + k \times y_{rand} \tag{8}$$

Here, $\hat{y}$ serves as the precise prediction value, y_{rand} provides the basis for interval prediction, and k acts as the interval control factor, determining the size of the interval and the prediction probability. $\hat{y}$, as the precise prediction value, is calculated using the following formula:

$$\hat{y}_{1,t}=\beta_0+\beta_1\times y_{1,t-1}+\beta_2\times y_{1,t-2}+\beta_3\times y_{1,t-3}+\beta_4\times y_{2,t-1}+\beta_5\times y_{2,t-2}+\beta_6\times y_{2,t-3}+\beta_7\times p \tag{9}$$

$$\hat{y}_{2,t}=\beta_8+\beta_9\times y_{1,t-1}+\beta_{10}\times y_{1,t-2}+\beta_{11}\times y_{1,t-3}+\beta_{12}\times y_{2,t-1}+\beta_{13}\times y_{2,t-2}+\beta_{14}\times y_{2,t-3}+\beta_{15}\times p \tag{10}$$

Here, $\hat{y}_{1,t}$ represents the medal count for ball sports in the t-th Olympic Games, and $\hat{y}_{2,t}$ represents the medal count for non-ball sports in the t Olympic Games. p is a binary variable (0 or 1), indicating whether the country is classified as a sports powerhouse, as introduced in the data preprocessing section. y_{rand}, serving as the basis for interval prediction, primarily reflects factors influencing the uncertainty of the prediction interval. These elements collectively contribute to the uncertainty in predicting each country's medal count:

$$y_{rand}=E_{leader}+E_{host}+E_{number}+E_{quin} \tag{11}$$

Next, the prediction process using multiple regression is introduced using $\hat{y}_{1,t}$ as an example:

Initialize Parameters:

- *Start with initial values for $\beta_0, \beta_1, \ldots, \beta_n$.*

Define the Cost Function:

- *Use Mean Squared Error (MSE):*

$$J(\beta) = \frac{1}{m}\sum_{i=1}^{m} \left(\left(\hat{y}_{1,t} - \left(\beta_0 + \sum_{j=1}^{3} \beta_j y_{1,t-j} + \sum_{j=4}^{6} \beta_j y_{2,t+3-j} z + \beta_7 p\right)\right)^2\right. \tag{12}$$

- *m: Number of data points.*
- *Minimize J(β) to find the optimal coefficients.*

Gradient Descent for Optimization:

- *Update coefficients iteratively:*

$$\beta_j := \beta_j - \alpha\frac{\partial J(\beta)}{\partial \beta_j}$$

- *α: Learning rate.*

Stop When Convergence is Reached:

- *Terminate the loop when changes in β_j are below a threshold.*

Algorithm: Multivariate Regression
Input: - Dataset with m samples: $(\hat{Y}, Y)$ - $\hat{Y}$: Independent variables (matrix of size $m \times 8$) - Y: Dependent variable (vector of size m) - Learning rate: alpha

- Number of iterations: *max_iters*

Output:

- Coefficients: beta (vector of size *n+1*

Steps:

1. Initialize beta = [0, 0,..., 0] (size *n+1*)
2. Add a column of one's toY for the intercept term
3. Repeat for *max_iters*:
   a. Calculate predictions: $\hat{Y}_p red = Y \times beta$
   b. Compute errors: error =$\hat{Y}\_pred - \hat{Y}$
   c. Calculate gradient: gradient = $(1/m) \times (Y^\wedge T \times error)$
   d. Update beta: beta = $beta - alpha \times gradient$
4. Return beta

If, for a country, the gold medal count for ball sports in this Olympic game increases by 1, the gold medal count for ball sports in the next future Olympic game will increase by approximately 0.729. Meanwhile, we observe a strong, positive correlation between $y_{1,t-2}$, and the total medal counts for ball sports, with a coefficient of 0.401. Also, it is worth noticing that the coefficient for the binary variable p, for both ball sports' gold counts and total medal counts, is 2.002, which highlights the role of powerhouse classification in medal predictions.

When the gold medal count for non-ball sports in this Olympic game increases by 1, the same country's gold medal count for ball sports in the next Olympic game will increase by approximately 0.975, nearly a 1:1 positive correlation. Meanwhile, the coefficient for the binary variable p (-9.559) suggests a surprisingly negative influence, which may indicate overfitting or a miscalibrated model. Meanwhile, the coefficient for p is significantly positive (33.241), highlighting powerhouse classification's role in total medal predictions for non-ball sports.

3.2.2 Random Prediction (Simulated Annealing Optimization)

After completing the construction of the y_{random}, the specific prediction steps of the prediction model in our first question are as follows. As we have already finished ball-sport and non-ball sport categorization and have made a preliminary estimation, our y_{random} will come into play by rounding down and up the formula to lower our MSE:

$$y_{pred} = \hat{y} + k \times y_{random} \tag{13}$$

In the simulated annealing algorithm (SA), the objective function E(s) incorporates the mean squared error decline and a penalty term, guiding the optimization process to balance prediction accuracy and interval size while iteratively searching for the optimal k value:

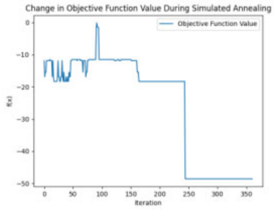
$$E(s) = maxMSE_{decline} - k^2 \tag{14}$$

$$MSE_{decline} = \frac{\sum_{i=1}^{m} (\hat{y} - y)^2}{m} \tag{15}$$

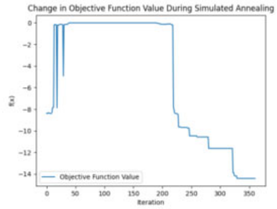
Pseudocode:

Algorithm: Simulated Annealing

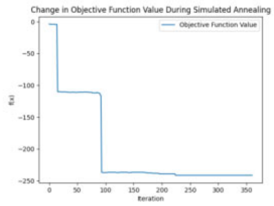
Let $k = k_0$
For $n = 0$ through n_{max} (exclusive):
$T \leftarrow$ temperature$(1 - (n+1)/n_{max})$
Pick a random neighbour$(k_{max}) \leftarrow$ neighbour(k)
If $P(E(k), E(k_{new}), T) \geq random(0,1)$:
$k \leftarrow k_{new}$
Output: the final state *k*



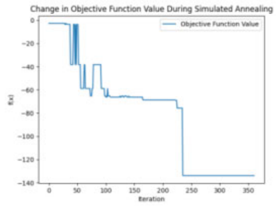
a. Ball-sport gold counts



b. Ball-sport total counts



c. Non-ball Sports gold counts



d. Non-ball sports total counts

Fig. 13 SA Results

The final k value for ball sports' gold medal predictions is 31, corresponding to an interval confidence level of 96.9%. Similarly, we apply this algorithm to the other three models and obtain results as shown below.

Table 1 K values, intervals

k	Values
Gold	31
Total	22.43
Interval confidence level	**Values**
Gold	0.9690
Total	0.9600

a. Ball sport

k	Values
Gold	34.0644
Total	29.9148
Interval confidence level	**Values**
Gold	0.9094
Total	0.9930

b. Non-ball sport

Next, we will make predictions for the top-performing countries in the 2024 Paris Olympics, keeping the ball and non-ball sports categorizations the same.

Table 2 Top 10 countries in the 2024 Paris Olympics

Rank	Country
1	United States
2	China
3	United Kingdom
4	Japan
5	Germany
6	Russia
7	France
8	Italy
9	South Korea
10	Brazil

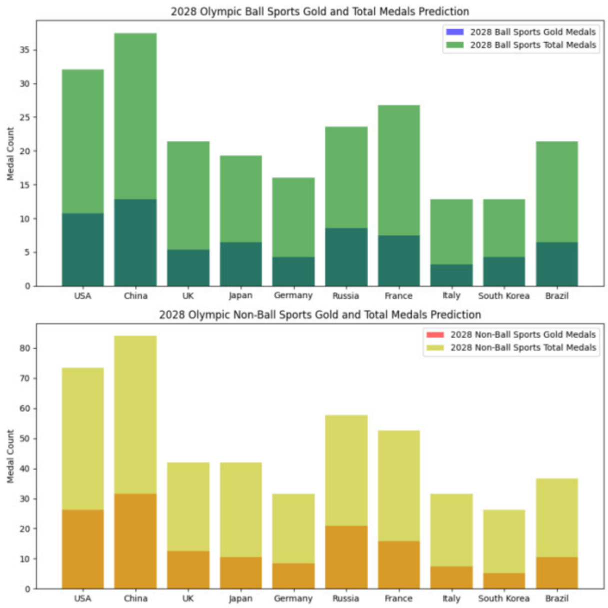


Fig. 14 Predictions

3.2.3 Prediction of Medal-Winning Countries in the Next Olympics

We build upon the previously established predictive model to make predictions of countries likely to win their first Olympic medal and calculate the probability of such predictions. For countries that have not yet won any medals, the deterministic component in our model is zero. That reflects that these countries have no prior achievements in medal-winning history. However, due to the influence of stochastic factors such as the host effect, the leadership effect (particularly prominent in developing sporting nations), and the candidate & incremental talent effects, these countries still hold significant potential for winning their first medal. The results show that the product of their k-value and the stochastic factor interval for the identified nations exceeds 1. Specifically, the probabilities for ball-sport and non-ball-sport predictions are 0.993 and 0.960, respectively. Taking the minimal value among the two, we conclude that the final prediction probability is approximately 96%.

4. The "Great Coach" Effect

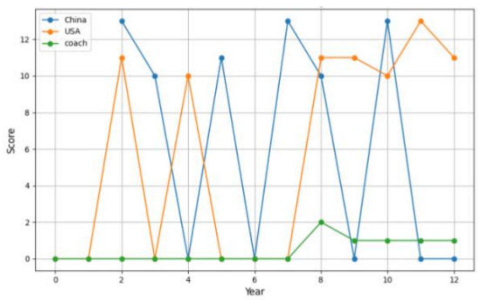


Fig. 15 Scores of China and USA by Year

The chart above compares the performance of the Chinese and U.S. women's volleyball teams, with significant changes between 2005 and 2008 (labeled as 5 and 8) can be observed from it. The Chinese team's performance declined during this period, while the U.S. team's performance improved notably. According to available information, Lang Ping served as the head coach of the U.S. women's volleyball team from 2005 to 2008. Under her leadership, the U.S. team achieved outstanding results, including a silver medal at the 2008 Beijing Olympics (Huang, 2016). In the chart, the period when Lang Ping coached the U.S. team is marked as "2", while the period when she coached the Chinese team is marked as "1," as shown by the green life, which together provides preliminary evidence supporting the impact of an "excellent coach effect."

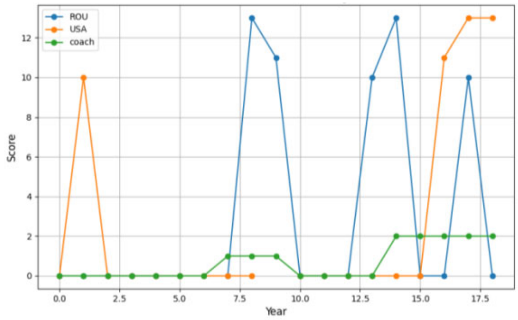


Fig. 16 Scores of ROU and USA by Year

Evidently, in the 1970s, the Romanian women's gymnastics team rose to prominence under her guidance, achieving remarkable success. Later, in 2001, the U.S. women's gymnastics team broke its long-standing medal drought, marking a significant breakthrough. Karolyi coached the Romanian squad during the 1970s and 1980s, helping them achieve numerous international victories, particularly at the 1976 and 1980 Olympic Games. In 1981, she began coaching the U.S. women's gymnastics team and became the U.S. national team coordinator in 2001, a position she held until her retirement in 2016 (USA Gymnastics). To further analyze, we calculated the Pearson correlation coefficient between the coaching years and the scores of the two countries and presented the results using heat maps.

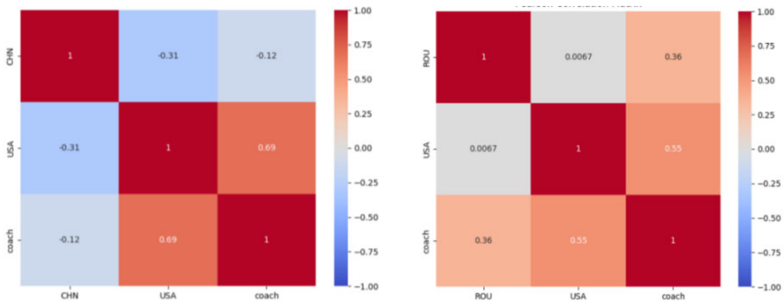


Fig. 17 Pearson Correlation Heat Maps: Lang Ping and Marta Karolyi

The left figure illustrates the correlation coefficients between Lang Ping's coaching tenure and the performance of China and the United States. In contrast, the correct figure represents the results for the second example. Both figures above illustrate the notable impact of exceptional coaches on scores for the analyzed countries. In the two highlighted cases, weaker teams experienced improvements of 0.69 and 0.55, both exceeding 0.5, indicating strong correlations. That demonstrates that outstanding coaches can lead to substantial progress in specific sports for their teams during a given Olympic cycle. Since the correlation coefficients for individual sports consistently exceed 0.5, the contribution of exceptional coaches to a country's total medal count can also be estimated at above 0.5.

5. The Original Insights on the Number of Olympic Medals

5.1 Categorization of Sports and Trends

The model identifies a degree of correlation among Olympic sports, though the strength of this correlation varies. By analyzing these relationships, this study shows that sports can be broadly categorized into "ball sports" and "non-ball sports." Sports exhibit similar patterns in medal-winning capacity and performance trends within each category, whereas minimal similarity exists across categories. This classification highlights the overarching characteristics of Olympic sports and has been integrated into the modeling process. However, further research is needed to explore whether this classification sufficiently reflects intrinsic connections among sports, such as technical demands, athletes' physical attributes, and training regimens. Other possible classifications—energy expenditure, team versus individual dynamics, or collaborative effort levels—might complement or refine these findings, offering more nuanced insights for resource allocation and strategic planning.

5.2 Dynamic Evolution of the Host Nation Effect

The study reveals a diminishing host nation effect as Olympic processes become more equitable. This trend reflects improvements in rules and organizational systems, which foster a more competitive and fair environment. Nonetheless, specific sports, particularly team events or those with subjective scoring, may still see host nations leverage their home-field advantages and resource access to influence outcomes. National Olympic committees can use these insights to strategically prioritize and prepare for specific events where host nation effects are likely to persist. Future research should investigate how event-specific characteristics, judging criteria, and audience dynamics impact the host nation effect to inform targeted interventions better.

References

[1] Yu, X., et al. Assessment and Adjustment of Approximate Inference Algorithms Using the Law of Total Variance[J]. Journal of Computational and Graphical Statistics, 2021, 30(4): 977-990.

[2] Howell, D. C. Statistical methods for psychology[M]. 7th ed. Wadsworth: Cengage Learning, 2010.

[3] Csurilla, G., Fertő, I. The less obvious effect of Olympics hosting on sporting performance[J]. Scientific Reports, 2023, 13: 819.

[4] Sacks, E. Why are some sports removed from the Olympic Games? it’s complicated - here’s why[EB/OL]. NBC Insider Official Site, 2024[2025-01-27].

[5] Huang, E. Eight years ago, China’s hero volleyball coach was a ‘traitor’ who coached the U.S. to victory[EB/OL]. Quartz, 2016[2025-01-27].

[6] USA Gymnastics. Coaching Team - Bela and Martha Karolyi[EB/OL]. USA Gymnastics,n.d.[2025-01-27].