Problem Chosen	2025	Team Control Number
A	**MCM/ICM** **Summary Sheet**	**2504496**

Enjoy a Cozy and Green Bath

Summary

abstract...

Keywords: Keyword one, Keyword two, Keyword three

Contents

1 Introduction

1.1 Background

The medal table of the 2024 Paris Olympics shows that the United States and China each won 40 gold medals and tied for the top spot, but the United States led with a total of 126 medals. The host country France ranked fifth in gold medals (16) and fourth in total medals (64). Dominica, Saint Lucia and other countries won their first Olympic medals, while 60 countries still have not broken through for any medals.

Figure 1: The medals of the 2024 Paris Olympics

1.2 Restatement and Analysis of the Problem

Based on the provided historical data-set of the Olympic Games from 1896 to 2024, we are employed to analyze and answer the following questions:

1. Develop a **prediction model** to forecast the number of medals each country will win in 2028, and identify countries that may progress or regress.

2. Provide **prediction intervals** and estimates of **uncertainty** and metrics to measure the model's performance.

3. Estimate the number of countries that will win their **first medal** and the probability of this happening.

4. Analyze the **relationship** between specific Olympic events (in terms of quantity and type) and the number of medals, explore which events are more important, and the impact of the host country's event selection strategy on the outcome.

5. Verify whether the **mobility of coaches** significantly enhances a country's performance in specific sports (such as Lang Ping and Bela Karolyi).

6. Quantify the contribution of **coaching effectiveness** to the number of medals, and recommend key sports for investment and expected returns for the three countries.

7. Extract the less-attended-to patterns from the model and provide strategic **suggestions** for the Olympic Committee.

For Task 1, we selected seven indicators and established an LSTM-based medal quantity prediction model, and provided interval predictions using Bayesian estimation. As for countries that have never won medals, we built an SVM-based "first medal breakthrough" prediction model based on the new events, the number of athletes, and historical participation trends.

1.3 Overview of Our Work

2 Assumptions and Justification

To simplify the problem and make it convenient for us to simulate real-life conditions, we make the following basic assumptions, each of which is properly justified.

- **1**. ...

- **2**. ...

3 List of Notations

Symbols	Description
$A_C, A_T, A_S, A_H(t)$	Set of country, years, all sports and host country in year t.
$n_{Gold}(t,i,j,k)$	Number of gold medals country i won in sport j at event k in year t.
$n_{Silver}(t,i,j,k)$	Number of silver medals country i won in sport j at event k in year t.
$n_{Bronze}(t,i,j,k)$	Number of bronze medals country i won in sport j at event k in year t.
$n_{Total}(t,i)$	Number of total medals country i won in year t.
$N_{participate}(t,i)$	Total number of athletes from country i in year t.
$N_{award}(t,i)$	Number of athletes who won medals from country i in year t.
$H(t,i)$	Host effect.
$y(t,i,j,k)$	Logical variable of event j of sport i from i in year t.

where we define the main parameters while specific value of those parameters will be found in the data-set attached.

4 Data Pre-processing

4.1 Outlier and Missing Value Handling

As the **1906 Intercalated Games** lacked the medal data of various countries and the competition results were not recognized by the International Olympic Committee, the data of 1906 is not taken into account.

In adition, **Skating** and **Ice Hockey** have been included in the Winter Olympics since 1920, so these two events are not within the scope of consideration. Otherwise, the "·" is replaced by the number 0.

It was noticed that **Jeu de Paume** and **Roque** sports in the **summerOly_programs.csv** do not have Codes. Upon researching information from https://en.wikipedia.org/wiki/Jeu_de_paume and https://en.wikipedia.org/wiki/Roque, it was found that only a few people are still engaged in these two sports, which have even not been held for 26 consecutive years in the Summer Olympics. Therefore, these two sports have been excluded.

5 Model 1: Prediction of Number of Medals for Medal-Winning Countries

5.1 Significance Analysis of Host Effect

Host Effect refers to the phenomenon where the host country or region performs more prominently in large-scale international events (such as the Olympic Games, the World Cup, etc.) due to its home field advantage. This phenomenon is usually reflected in a significant increase in the host country's medal count, competition results, or overall performance.

To verify the significance of the host effect, we employed a paired samples t-test. Firstly, we selected the medal count of the host country in each year, denoted as MT_t, as the first sample. Secondly, to eliminate the influence of the overall growth trend in medal counts, we took the average of the medal counts obtained by the host country in the two consecutive Olympic Games as the second sample, shown as (1),

$$MT_t^s = \frac{MT_{t-1} + MT_{t+1}}{2} \tag{1}$$

where $t \in A_T^H = A_T \backslash \{1896, 2024\}$, $i \in A_C$. Then $\{MT_t, MT_t^s\}$ constitutes paired data with a sample size of 30. Define $d_t = MT_t - MT_t^s$, and assume that

$$H_0 : d_{1900} = \cdots = d_{2020}, \quad vs \quad H_1 : \exists t_1, t_2 \in A_T^H \text{ s.t. } d_{t_1} \neq d_{t_2}.$$

Select the t-test statistic as

$$t = \frac{\bar{d}}{s_d/30} \sim T(29) \tag{2}$$

where $\bar{d} = \frac{1}{30} \sum_{t \in A_T^H} d_t$ is the mean of paired samples, and $s_d = \frac{1}{29} \sum_{t \in A_T^H} \left(d_t - \bar{d} \right)^2$ is the sample variance of the differences of paired data,

For a given significance level α, the rejection domain for the hypothesis test is

$$W_\alpha = \left\{ t \geq T_{1-\alpha}(29) \right\} \tag{3}$$

By following the described procedure, the results of the t-test were obtained and are summarized in Table **??**.

5.2 Index Analysis

5.2.1 Host effect

Define Logical Variable $H(t, i)$ as equation (4),

$$H(t, i) = \begin{cases} 1, & \text{Country } i \text{ is host in year } t, \\ 0, & \text{others.} \end{cases} \tag{4}$$

where $t \in A_T$, $i \in A_C$.

5.2.2 Dominant Event

A **Dominant Event** is defined as a discipline in which a National Olympic Committee (NOC) consistently achieves a high medal yield, contributing significantly to its overall medal tally. The dominance is quantified by the **medal ratio**, which is the proportion of medals earned in a specific event relative to the NOC's total medal count during a given Olympic Games.

$$P(t,i) = \frac{N_{award}(t,i)}{N_{participate}(t,i)}$$

5.2.3 Other indexes

5.3 LSTM Model

6 Model 2: Prediction of Maiden Medal for Medal-Less Countries

7 Task 2: xxx

8 Task 3: xxx

9 Task 4: xxx

10 Sensitivity Analysis

11 Strength and Weakness

11.1 Strength

11.2 Weakness

12 Further Discussion

Memo

Enjoy Your Bath Time!

Sincerely yours,

Your friends

Reference

Appendices

Appendix A   First appendix

Appendix B   Second appendix

Report on Use of AI

1. OpenAI ChatGPT (Nov 5, 2023 version, ChatGPT-4,)

   Query1: <insert the exact wording you input into the AI tool>
   Output: <insert the complete output from the AI tool>

2. OpenAI ChatGPT (Nov 5, 2023 version, ChatGPT-4,)

   Query1: <insert the exact wording you input into the AI tool>
   Output: <insert the complete output from the AI tool>