**2025 MCM**

**Problem C: Models for Olympic Medal Tables**

**问题 C：奥运奖牌榜模型**



|  |
| --- |
|  |

In addition to watching individual **events** during the latest summer Olympic Games in Paris,

2024, fans followed the overall “medal table” for each country. The final results (**Table 1**)

showed the United States with the most total medals (126), and China and the United States tied at the top of the standings for the first place Gold medals (40). The host country, France, was

5th in the Gold medal count (16) standings but 4th in terms of total medal count while Great Britain, 7th with 14 Gold medals finished 3rd in total medals.

在最近一届 2024 年巴黎夏季奥运会期间，除了观看单项比赛外，粉丝们还关注各国的总 “奖牌榜”。最终结果（表 1）显示，美国获得的奖牌总数最多（126 枚），中国和美国并列金牌榜首位（40 枚）。东道主法国在金牌数（16 枚）排行榜上名列第五，但在奖牌总数上名列第四，而英国以 14 枚金牌名列第七，在奖牌总数上名列第三。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Gold** | **Silver** | **Bronze** | **Total** |
| **United States** | 40 | 44 | 42 | 126 |
| **China** | 40 | 27 | 24 | 91 |
| **Japan** | 20 | 12 | 13 | 45 |
| **Australia** | 18 | 19 | 16 | 53 |
| **France** | 16 | 26 | 22 | 64 |
| **Netherlands** | 15 | 7 | 12 | 34 |
| **Great Britain** | 14 | 22 | 29 | 65 |

**Table 1:** Paris Olympics (2024) Final Medal Table – Gold Medal Top 7 Countries**[1]**

The standings at the top of the table are always watched closely, but the medal counts for other

countries are often just as valued. For example, Albania (2 medals), Cabo Verde, Dominica, and Saint Lucia (2 medals) won their nations' first Olympic medals at the Paris games. Dominica and Saint Lucia also each earned a Gold medal at these games. More than 60 countries have still yet to win an Olympic medal.

Predictions of the final medal counts are commonly made, but typically not based on historical medal counts but closer to the start of an upcoming Olympic games when current athletes

scheduled to compete are known (for example:

[https://www.nielsen.com/news-center/2024/virtual-medal-table-forecast/)](https://www.nielsen.com/news-center/2024/virtual-medal-table-forecast/).

排名靠前的国家总是受到密切关注，但其他国家的奖牌数往往也同样受到重视。例如，阿尔巴尼亚（2 枚奖牌）、佛得角、多米尼克和圣卢西亚（2 枚奖牌）在巴黎奥运会上赢得了本国的首枚奥运奖牌。多米尼克和圣卢西亚还在这些比赛中各获得一枚金牌。目前仍有 60 多个国家尚未获得奥运奖牌。

对最终奖牌数的预测很常见，但通常不是基于历史奖牌数，而是在临近即将举行的奥运会开幕时，当已知目前计划参赛的运动员时进行预测

(例如：<https://www.nielsen.com/news-center/2024/virtual-medal-table-forecast/）。>

Data is provided of medal tables for all summer Olympic games, host countries, as well as the

number of Olympic events at each games broken down by **sport** for all summer Olympic games played. Additionally, data for all individual Olympic competitors with their sport and result

(medal type, or no medal) is provided. Your models and data analysis must ONLY use the

provided data sets. You may use additional resources to provide background and context or help with interpreting results (be sure to document the sources). Specifically, use the provided data to:

提供的数据包括所有夏季奥运会的奖牌榜、主办国，以及所有夏季奥运会按运动项目分列的每场比赛的奥运项目数量。此外，还提供了所有奥运选手的个人数据，包括他们的运动项目和成绩（奖牌类型或无奖牌）。您的模型和数据分析必须只使用所提供的数据集。您可以使用其他资源来提供背景情况或帮助解释结果（请务必记录来源）。具体来说，使用所提供的数据来：

• Develop a model for medal counts for each country (for Gold and total medals at a

minimum). Include estimates of the uncertainty/precision of your model predictions and measures of how well model performs.

o Based on your model, what are your projections for the medal table in the Los Angeles, USA summer Olympics in 2028? Include prediction intervals for all results. Which countries do you believe are most likely to improve? Which will do worse than in 2024?

o Your model should include countries that have yet to earn medals; what is your projection for how many will earn their first medal in the next Olympics? What sort of odds do you give to this estimate?

o Your model should also consider the events (number and types) at a given

Olympics. Explore the relationship between the events and how many medals

countries earn. What sports are most important for various countries? Why? How do the events chosen by the home country impact results?

• 为每个国家的奖牌数建立一个模型（至少包括金牌数和奖牌总数）。包括对模型预测的不确定性/精确度的估计，以及对模型性能的衡量。

o 根据您的模型，您对 2028 年美国洛杉矶夏季奥运会奖牌榜有何预测？包括所有结果的预测区间。您认为哪些国家的成绩最有可能提高？哪些国家的成绩会比 2024 年差？

o 您的模型应包括尚未获得奖牌的国家；您预测有多少国家将在下届奥运会上获得第一枚奖牌？您对这一估计给出的概率是多少？

o 您的模型还应考虑特定奥运会的项目（数量和类型）。探索项目与国家获得奖牌数量之间的关系。哪些项目对不同国家最重要？本国选择的项目对结果有何影响？

|  |
| --- |
| 问题要求：  建立模型，能够实现预测 2028 年美国洛杉矶夏季奥运会奖牌榜的预测   * 探索项目与国家获得奖牌数量之间的关系   可以先做一些探索性分析，可以使用相关性分析或线性回归分析项目的数量对国家获得奖牌数量之间的关系，采用单因素方差分析项目的类型对国家获得奖牌数量之间的关系；  也可以在建模后，采用指标重要度来分析2者之间的影响   * 需考虑奥运会的项目（数量和类型），预测各国2028年金牌数和奖牌总数   在纳入以上指标的情况，可以通过两种机器学习建模来解决：   1. ovo模式，也就是单独预测，单独与奖牌总量、3种奖牌数分别建立4个模型进行预测；模型上可以采用机器学习回归预测模型，例如决策树、xgboost、lgbm、lstm等等 2. 序列预测，多输入与多输出，设置多个输出作为模型预测，模型上可以采用深度学习回归预测模型，例如lstm或seq2seq模型等等   需要预测得准确，可以从下面三个方面做模型糅合：   * 主办国效应（Host Effect）：主办国往往在奖牌数上表现更佳，这需要纳入模型。   可以直接用哑变量来标识   * 运动项目相关性：分析不同项目对各国奖牌数的影响。   可以对历届各国获奖的金牌运动类型进行变量化，标识获奖率（参加次数/获奖概率）   * 历史奖牌趋势（Trend）：利用过去几届奥运会的数据，提取国家的奖牌数增长/衰退趋势。   有现成数据 |

• Athletes may compete for different countries, but it is not a simple matter for them to

change due to citizenship requirements. Coaches, however, can easily move from one

country to another as they do not need to be citizens to coach. There is, therefore, the

possibility of a “great coach” effect. Two possible examples of this include Lang Ping**[2]**, who coached volleyball teams from both the U.S. and China to championships, and the

sometimes-controversial gymnastics coach, Béla Károlyi**[3]**, who coached Romania and then the U.S. women’s teams with great success. Examine the data for evidence of

changes that might be due to a “great coach” effect. How much do you estimate such an effect contributes to medal counts? Choose three countries and identify sports where they should consider investing in a “great” coach and estimate that impact.

• 运动员可以代表不同的国家参加比赛，但由于公民身份的要求，他们并不容易改变。但是，教练员却可以很容易地从一个国家转到另一个国家，因为他们不需要公民身份来执教。教练员却可以很容易地从一个国家转到另一个国家，因为他们不需要成为公民就可以执教。因此，有可能产生 “伟大教练 ”效应。郎平[2] 就是两个可能的例子，他曾执教中美两国的排球队都获得过冠军。

贝拉-卡洛里[Béla Károlyi][3]，他先后执教罗马尼亚和美国女子体操队并取得巨大成功。研究数据，寻找可能由 “伟大教练 ”效应引起的变化的证据。您估计这种效应对奖牌数的影响有多大？选择三个国家，确定它们应考虑投资 “优秀 ”教练的体育项目，并估计其影响。

|  |
| --- |
| 分2步：   1. 证明由 “伟大教练 ”效应引起的变化的证据，估计这种效应对奖牌数的影响有多大？   针对第一个问题，则需要收集数据，结合外部来源，收集国际知名教练执教历史（如郎平和贝拉-卡洛里），将教练执教国家与该国奖牌数和项目成绩的变化趋势进行差异性分析。采用的模型例如独立样本T检验或者did双重差分分析，需要注意，收集数据时，最好侧重三个项目，也就是我们第二问需要分析的项目   1. 选择三个国家，确定它们应考虑投资 “优秀 ”教练的体育项目，并估计其影响。   确定它们应考虑投资 “优秀 ”教练的体育项目应该是他们参与人数最多的项目（当然你可以选择其他口径，能自圆其说即可），选择3个国家，估计其影响，这里需要结合上面的优秀知名教练的影响提升比例，作为提升后的效果即可，然后拿那个比例套入该国家的参赛人员数，来估计他的影响 |

• What other original insight(s) about Olympic medal counts does your model reveal? Explain how these insight(s) can inform country Olympic committees.

• 您的模型还揭示了哪些有关奥运奖牌数的独到见解？解释这些见解如何为各国奥委会提供信息。

Your PDF solution of no more than 25 total pages should include:

• One-page Summary Sheet.

• Table of Contents.

• Your complete solution.

• References list.

• [AI Use Report](https://www.contest.comap.com/undergraduate/contests/mcm/flyer/Contest_AI_Policy.pdf)(If used does not count toward the 25-page limit.)

您的 PDF 解决方案总页数不超过 25 页，其中应包括

- 一页摘要表。

- 目录。

- 完整的解决方案。

- 参考文献列表。

- 人工智能使用报告（如已使用，则不计入 25 页限制。）

**Note:** There is no specific required minimum page length for a complete MCM submission. You may use up to 25 total pages for all your solution work and any additional information you want to include (for example: drawings, diagrams, calculations, tables). Partial solutions are accepted. We permit the careful use of AI such as ChatGPT, although it is not necessary to create a solution to this problem. If you choose to utilize a generative AI, you must follow the[COMAP AI use](https://www.contest.comap.com/undergraduate/contests/mcm/flyer/Contest_AI_Policy.pdf)

[policy.](https://www.contest.comap.com/undergraduate/contests/mcm/flyer/Contest_AI_Policy.pdf) This will result in an additional AI use report that you must add to the end of your PDF solution file and does not count toward the 25 total page limit for your solution.

[NEW MCM/ICM: Online Submission Process](https://www.contest.comap.com/undergraduate/contests/mcm/flyer/MCM-ICM_SubProcess.pdf)

The purpose of this article is to assist and guide students and advisors participating in HiMCM/MidMCM. In the article, COMAP, provides information about the new online submission process using the new online submission page

[https://forms.comap.org/241335097294056.](https://forms.comap.org/241335097294056) You will need your team's control number, advisor id number and your problem choice to complete your submission.

**Data Files**

**2025\_Problem\_C\_Data.zip: This zip file contains all 5 of the data files listed below.**

[•](#bookmark1) *[data](#bookmark1)*[\_](#bookmark1)*[dictionary.csv](#bookmark1)* [– database descriptions with examples](#bookmark1)

[•](#bookmark2) *[summerOly](#bookmark2)*[\_](#bookmark2)*[athletes.csv](#bookmark2)*[– all competitors with their sport, year, and result (medal type or](#bookmark2)

[none)](#bookmark3)

[•](#bookmark4) *[summerOly](#bookmark4)*[\_](#bookmark4)*[medal](#bookmark4)*[\_](#bookmark4)*[counts.csv](#bookmark4)* [– complete country medal count tables for all summer](#bookmark4)

[Olympics from 1896 to 2024](#bookmark5)

[•](#bookmark6) *[summerOly](#bookmark6)*[\_](#bookmark6)*[hosts.csv](#bookmark6)* [– list of host country for all summer Olympics from 1896 to 2032](#bookmark6)

[•](#bookmark7) *[summerOly](#bookmark7)*[\_](#bookmark7)*[programs.csv](#bookmark7)*[– counts of number of events by sport/](#bookmark7)**[discipline](#bookmark7)** [and total for](#bookmark7)

[all summer Olympics from 1896 to 2032](#bookmark8)

2025\_Problem\_C\_Data.zip: 此压缩文件包含下面列出的全部 5 个数据文件。

- data\_dictionary.csv - 包含示例的数据库说明

- summerOly\_athletes.csv - 所有选手的运动项目、年份和成绩（奖牌类型或

无）

- summerOly\_medal\_counts.csv - 1896 年至 2024 年所有夏季奥运会的完整国家奖牌总数表

- summerOly\_hosts.csv - 1896 年至 2032 年所有夏季奥运会的主办国列表

- summerOly\_programs.csv - 1896 年至 2032 年所有夏季奥运会按运动/项目和总数分列的项目数量统计表

Data, such as country designations, are recorded by the **International Olympic Committee**

**(IOC)** (on their Olympics.com website) at the time of a given Olympics. Thus, designations may change in the data set. As with all data, there may be recording anomalies. Note, for example, in the athlete’s data set in some cases for sports like tennis, table tennis, beach volleyball, the

“Team” includes more detail than just the country. For example, Germany-1 would be the first of two beach volleyball teams from Germany in the 2000 Olympics. Decisions and assumptions

about how to handle the data are an important part of the modeling process.

国家名称等数据由国际奥委会（IOC）（在其 Olympics.com 网站上）在特定奥运会举行时记录。因此，数据集中的名称可能会发生变化。与所有数据一样，可能存在记录异常的情况。例如，在网球、乒乓球、沙滩排球等运动项目的运动员数据集中，“团队 ”包含了比国家更多的细节。例如，“Germany-1 ”指的是 2000 年奥运会德国沙滩排球两支队伍中的第一支。如何处理数据的决定和假设是建模过程的重要组成部分。

**Glossary**

**International Olympic Committee (IOC)**: is the international, non-governmental, sports

governing body of the Olympic Games and the Olympic Movement. The IOC is best known as the organization responsible for organizing the Summer and Winter Olympics.

**Programme**: of the Olympic Games is the programme of all sports competitions established by the IOC for each edition of the Olympic Games.

术语表

**国际奥林匹克委员会（IOC）：**是奥运会和奥林匹克运动的国际性非政府体育管理机构。国际奥委会最著名的活动是负责组织夏季和冬季奥运会。

**奥运会项目表：**是国际奥委会为每届奥运会制定的所有体育比赛的项目表。

**SDE**: **Sport, Discipline,** or **Event**

**Sport**: The IOC defines an Olympic sport as a discipline that is governed by a single

international sports federation (IF). A single sport may contain one or more disciplines, each of which is the focus of one or more events.

**Discipline**: A branch of a sport that includes one or more events.

**Event**: A competition within a discipline that results in a ranking and awards (e.g. medals).

Example of the relationship between **sport**, **discipline**, and **event** in the Olympic programme from the 2024 Paris Olympics:

● World Aquatics is the IF that governs the sport of aquatics

● Within the sport of aquatics are multiple disciplines – artistic swimming, diving, marathon swimming, swimming, and water polo.

● Within the discipline of diving are eight medal events:

○ Individual 3m springboard - men & women

○ Individual 10m platform - men & women

○ Synchronized 3m springboard - men & women

○ Synchronized 10m platform - men & women

SDE： 运动、学科或赛事

运动： 国际奥委会将奥林匹克运动定义为由一个国际单项体育联合会（IF）管理的学科。一个运动项目可包含一个或多个学科，每个学科又是一个或多个赛事的重点。

学科： 一个运动项目的分支，包括一个或多个项目。

赛事： 学科内的比赛，产生排名和奖项（如奖牌）。

2024 年巴黎奥运会奥运项目中体育、学科和赛事之间的关系示例：

● 世界水上运动联合会是管理水上运动的国际联合会。

● 水上运动包含多个项目--艺术游泳、跳水、马拉松游泳、游泳和水球。

● 跳水项目中有八个奖牌项目：

○ 个人 3 米跳板--男子和女子

○ 个人 10 米跳台--男子和女子

○ 同步 3 米跳板 - 男子和女子

○ 同步 10 米跳台 - 男子和女子

**References**

**[1]** Olympics.com,<https://olympics.com/en/paris-2024/medals>

**[2]** Olympics.com Biography, Lang Ping,<https://olympics.com/en/athletes/ping-lang>

**[3]** USA Gymnastics Hall of Fame,

<https://usagym.org/halloffame/inductee/coaching-team-bela-martha-karolyi/>

**Disclaimer**

COMAP is a non-profit organization dedicated to improving mathematics education with an

emphasis on increasing student proficiency in mathematical modeling. This contest problem

references material from the International Olympic Committee (IOC). We acknowledge and

respect the IOC’s ownership of this material, and it is used here solely for educational, non-

commercial purposes to enrich learning experiences for participants. This content is not endorsed by or affiliated with the IOC.