|  |  |  |
| --- | --- | --- |
| **Problem Chosen** ABCDEF | **2025 MCM/ICM Summary Sheet** | **Team Control Number** 1111111 |

**Title**（此处应写论文标题）

**Summary**

* 美赛论文摘要的英文一般用Summary，摘要最好在本页完成，尽量不要超一页！

**标题**为16号Times New Roman字体加粗并居中

**摘要**为14号Times New Roman字体加粗并居中

**正文**为12号Times New Roman字体并两端对齐

**关键词**为12号Times New Roman字体加粗并居左

* **行间距一般为1倍行距，为控制在一页可适当调整，如设置为1倍行距或1.5倍行距等。**
* 首页一般不需要页眉和页码。
* “1111111”修改为自己的控制编号（Team Control Number），“ABCDEF”改为自己的选题题号（A/B/C/D/E/F）

使用此模板可以开始输入电子报告的第一页（摘要页）。这个模板使用了一个12点倍的新罗马字体。将您的论文以Adobe PDF电子文件（例如1111111.pdf）的形式提交，以英文输入，可读字体至少为12点类型。 不要在本页或任何页面上包括您的学校、顾问或团队成员的姓名。 论文必须在25页的限制范围内，包括正文、目录和附录等。 请务必更改上面的控制编号和问题选择。

**Key words:** 三到五个关键词

注：红色字为解释说明部分，使用时应全部删除或换成黑色字。

**Contents**

[I. Introduction 3](#_Toc471658476)

[1.1 Background 3](#_Toc471658477)

[1.2 Our works 3](#_Toc471658478)

[II. The Description of the Problem 4](#_Toc471658479)

[2.1 Problem statement 4](#_Toc471658480)

[2.2 Analysis of Specific Issues 4](#_Toc471658481)

[2.2.1 Analysis of Problem 1 4](#_Toc471658482)

[2.2.2 Analysis of Problem 2 4](#_Toc471658483)

[2.2.3 Analysis of Problem 3 4](#_Toc471658484)

[2.2.4 Analysis of Problem 4 5](#_Toc471658485)

[III. Basic assumption 5](#_Toc471658486)

[IV. Glossary & Symbols 5](#_Toc471658487)

[4.1 Glossary 5](#_Toc471658488)

[4.2 Symbols 6](#_Toc471658489)

[V. Models 6](#_Toc471658490)

[5.1 Analysis and Solving of Question One 6](#_Toc471658491)

[5.1.1 Model Preparation 6](#_Toc471658492)

[5.1.2 Model Establishment 7](#_Toc471658493)

[5.1.3 Results 8](#_Toc471658494)

[5.1.4 Analysis of the Result 8](#_Toc471658495)

[5.2 Analysis and Solving of Question Two 8](#_Toc471658496)

[5.2.1 Model Preparation 8](#_Toc471658497)

[5.2.2 Model Establishment 9](#_Toc471658498)

[5.2.3 Results 10](#_Toc471658499)

[5.2.4 Analysis of the Result 10](#_Toc471658500)

[5.3 Analysis and Solving of Question Three 10](#_Toc471658501)

[5.3.1 Model Preparation 10](#_Toc471658502)

[5.3.2 Model Establishment 12](#_Toc471658503)

[5.3.3 Results 12](#_Toc471658504)

[5.3.4 Analysis of the Result 13](#_Toc471658505)

[5.4 Analysis and Solving of Question Four 13](#_Toc471658506)

[5.4.1 Model Preparation 13](#_Toc471658507)

[5.4.2 Model Establishment 14](#_Toc471658508)

[5.4.3 Results 14](#_Toc471658509)

[5.4.4 Analysis of the Result 15](#_Toc471658510)

[VI. Error Analysis and Sensitivity Analysis 15](#_Toc471658511)

[6.1 Error Analysis 15](#_Toc471658512)

[6.1.1 Error Analysis of Model One 15](#_Toc471658513)

[6.1.2 Error Analysis of Model Two 15](#_Toc471658514)

[6.1.3 Error Analysis of Model Three 16](#_Toc471658515)

[6.2 Sensitivity Analysis 16](#_Toc471658516)

[6.2.1 Sensitivity Analysis of Model One 16](#_Toc471658517)

[6.2.2 Sensitivity Analysis of Model Two 16](#_Toc471658518)

[6.2.2 Sensitivity Analysis of Model Three 16](#_Toc471658519)

[VII. Evaluation and Promotion of Model 17](#_Toc471658520)

[7.1 Strength and Weakness 17](#_Toc471658521)

[7.1.1 Strength 17](#_Toc471658522)

[7.1.2 Weakness: 17](#_Toc471658523)

[7.2 Promotion 17](#_Toc471658524)

[Ⅷ. Conclusions 17](#_Toc471658525)

[8.1 Conclusions of the problem 17](#_Toc471658526)

[8.2 Methods used in our models 18](#_Toc471658527)

[I X. References 18](#_Toc471658528)

[X. Appendix 18](#_Toc471658529)

[10.1 Appendix One 18](#_Toc471658530)

[10.2 Appendix Two 19](#_Toc471658531)

从目录开始有页眉和页码，把“0000000”改为自己的控制编号，页码会自动调整可以不用编辑。

# I. Introduction

## Background

奥林匹克运动会作为全球规模最大的综合性体育赛事，每届夏季奥运会不仅展示了各国顶尖运动员的竞技水平，也反映了各国在体育领域的综合实力。2024年在巴黎举办的夏季奥运会上，美国以总奖牌数126枚位居榜首，中国和美国在金牌数上并列第一，各获得40枚金牌。东道主法国在金牌榜上排名第五，获得16枚金牌，但在总奖牌数上排名第四，达到64枚。英国以14枚金牌排名第七，但在总奖牌数上却位居第三，共赢得65枚奖牌。此外，阿尔巴尼亚、佛得角、多米尼加和圣卢西亚等国家在本届奥运会上首次摘得奖牌，其中多米尼加和圣卢西亚还分别获得了金牌。这些成绩不仅激励了各国体育发展的积极性，也为未来奥运奖牌预测提供了宝贵的数据支持。

预测奥运会的最终奖牌数目一直是体育分析的重要内容，然而大多数预测通常基于历史奖牌数据，或者在奥运会即将开始时根据现有运动员的参赛情况进行。随着数据分析技术的进步，利用更为详尽和全面的数据集进行奖牌预测成为可能。本研究旨在基于历届夏季奥运会的奖牌数据、主办国信息以及各届奥运会的比赛项目数量，构建一个精准的奖牌预测模型，以预测2028年洛杉矶夏季奥运会的奖牌分布情况，并探讨影响各国奖牌数变化的关键因素。



## Our works

在本研究中，首先针对奖牌数预测构建了一个多变量回归模型。

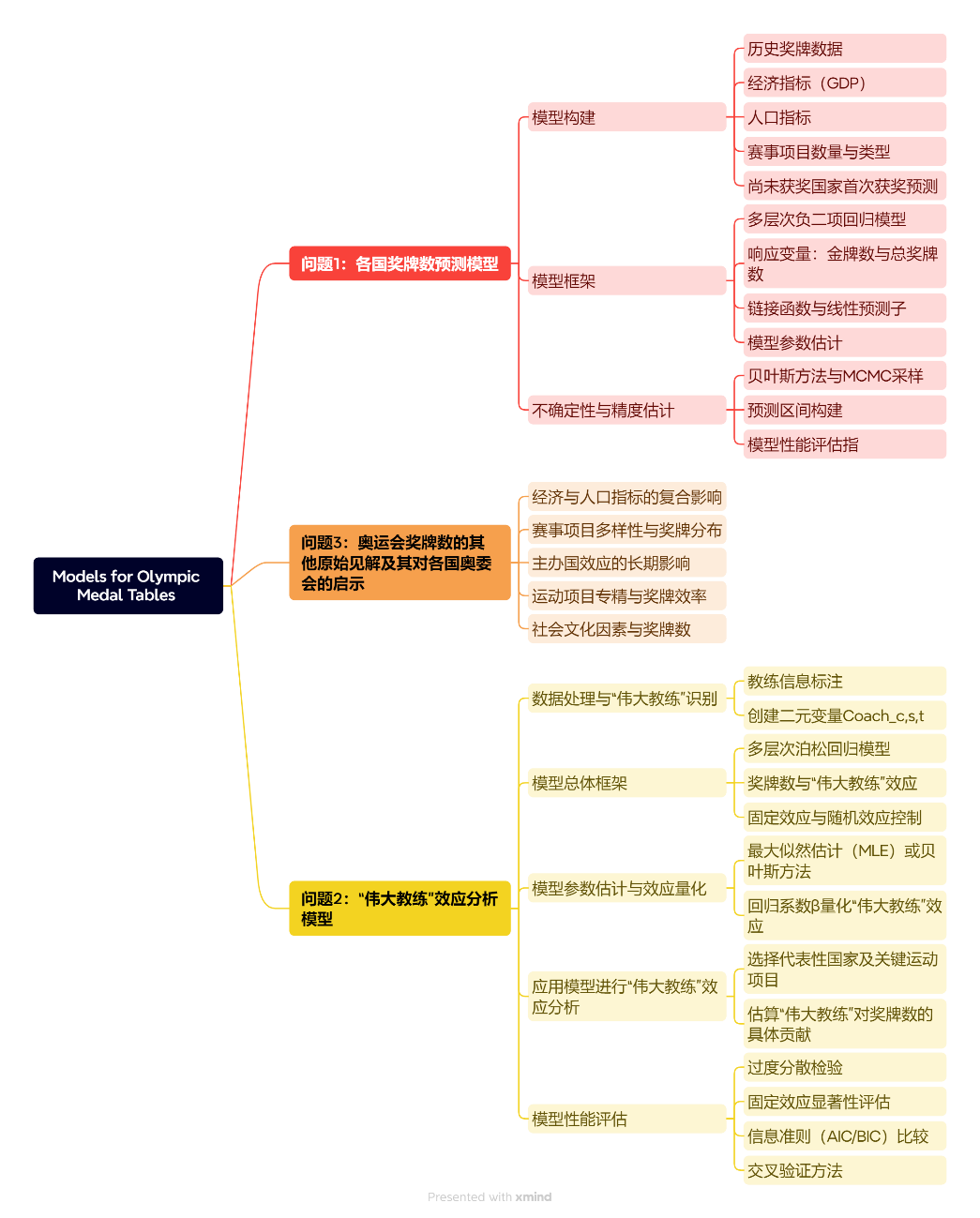
任务一的核心是该模型综合考虑了历届夏季奥运会的历史奖牌数据、主办国效应、各国在不同体育项目中的优势表现以及新增项目对奖牌分布的潜在影响。通过对这些变量的深入分析，不仅能够预测各国在2028年洛杉矶夏季奥运会上的金牌和总奖牌数，还能够为模型的不确定性和预测精度提供量化估计。此外，还通过预测区间的设定，评估了各国奖牌数变化的可能范围，识别出在未来奥运会上有望提升或可能表现不佳的国家。

在任务二中，重点研究了“伟大教练”效应对奥运奖牌数的影响。通过分析教练员跨国执教带来的成绩提升现象，评估了顶尖教练在提升特定国家某些体育项目奖牌数中的作用。具体而言，选取了三个具有代表性的国家，识别出这些国家在某些关键体育项目中通过引进优秀教练可以显著提升奖牌数的领域。基于历史数据和教练员执教前后的成绩变化，估算了“伟大教练”效应对奖牌数的具体贡献，并为这些国家未来在相关体育项目中的投资决策提供了数据支持和策略建议。

任务三则聚焦于模型揭示的其他奥运奖牌分布的独特见解。通过对模型结果的深入解读，发现了一些影响奖牌分布的关键因素，如主办国的项目设置对其奖牌数的显著影响、不同体育项目的竞争激烈程度以及新兴体育项目对奖牌分布的潜在重塑作用。这些发现不仅丰富了对奥运奖牌分布机制的理解，也为各国奥委会在制定体育发展战略时提供了有价值的参考。例如，某些国家在特定体育项目上的优势可以通过增加相关训练投入和资源配置得到进一步强化，从而在未来奥运会上取得更好的成绩。

# II. The Description of the Problem

通过对三个核心问题的深入分析，本研究全面构建并验证了各国在2028年洛杉矶夏季奥运会中的奖牌预测模型，系统评估了“伟大教练”效应对奖牌数的显著影响，并挖掘了影响奥运奖牌分布的多重因素。这一综合性的研究不仅为奖牌数的精准预测提供了科学依据，还揭示了教练资源配置和赛事项目多样性在提升国家竞争力中的关键作用。此外，模型所揭示的经济与人口复合影响、主办国效应及社会文化因素等见解，为各国奥委会制定长期体育发展战略、优化资源分配和提升训练体系提供了实证支持。具体问题思路如下图所示：





## 问题1：各国奖牌数预测模型

为了有效预测各国在2028年洛杉矶夏季奥运会中的金牌数和总奖牌数，我们构建了一个综合性的预测模型。该模型综合考虑了历届夏季奥运会的历史奖牌数据、各国的经济和人口指标、以及赛事项目的数量和类型等多重因素。通过分析这些变量之间的关系，我们能够估计每个国家在未来奥运会中的奖牌表现。此外，模型还特别关注那些尚未获得奖牌的国家，预测其在未来奥运会上首次夺牌的可能性。这一预测不仅依赖于国家的基本经济和人口状况，还结合了各国在不同体育项目中的历史优势和潜在发展空间，从而提供更为精准和全面的奖牌数预测。

## 问题2：“伟大教练”效应分析模型

在奥运会的奖牌竞争中，教练的作用不可忽视。为了量化“伟大教练”对各国奖牌数的影响，我们构建了一个专门的分析模型。该模型通过识别和评估顶尖教练在不同国家和运动项目中的执教经历，分析他们对运动队成绩的具体贡献。通过对比有“伟大教练”执教和无“伟大教练”执教期间的奖牌数变化，我们能够估算出优秀教练对提升国家奖牌数的具体效应。进一步地，我们选择了三个具有代表性的国家，结合其关键运动项目，评估投资“伟大教练”可能带来的潜在影响，从而为各国奥委会在教练资源配置和引进策略上提供科学依据。

## 问题3：奥运会奖牌数的其他原始见解及其对各国奥委会的启示

除了奖牌数的预测和“伟大教练”效应的分析外，我们的模型还揭示了一些关于奥运会奖牌分布的独特见解。这些见解包括经济与人口指标的复合影响、赛事项目的多样性对奖牌分布的影响、以及主办国效应的长期影响等。例如，经济发达和人口众多的国家在奖牌竞争中往往具有优势，但这种优势需要通过有效的资源配置和训练体系来实现。此外，赛事项目的多样性不仅影响总奖牌数，还决定了各国在特定项目上的竞争力。主办国效应则表明，作为主办国的国家在举办奥运会期间通常会获得更多奖牌，这一效应可能源自于基础设施的改善和运动员训练水平的提升。通过这些原始见解，各国奥委会可以更好地制定体育发展战略，优化资源投入，提升在国际赛事中的竞争力。

# III. Basic assumption

* **奖牌数的独立性与分布假设**

假设各国在不同届奥运会中的奖牌数相互独立，且金牌数服从负二项分布，以处理数据中的过度分散现象。这一假设确保模型能够准确反映实际数据的统计特性，提高预测的可靠性。

* **线性关系与链接函数的适用性**

假设各自变量（如GDP、人口规模等）与响应变量（金牌数和总奖牌数）之间存在线性关系，并采用对数链接函数将期望奖牌数与自变量的线性组合关联。这保证了模型的解释性和计算的简便性。

* **随机效应的正态分布假设**

假设模型中的国家和届次随机效应服从正态分布。这一假设允许模型捕捉到国家间及不同届次之间的异质性，控制不可观测的固定差异，从而提升模型的稳健性和泛化能力。

* **变量的充分性与无遗漏**

假设所选取的自变量能够充分解释奖牌数的变化，且模型中不存在重要的遗漏变量。包括经济指标、人口规模、历史奖牌数、是否为主办国等关键因素，以减少模型偏误，提升预测准确性。

* **教练效应的稳定性与可测性**

假设“伟大教练”的效应在不同国家和运动项目中具有一定的稳定性，并能够通过模型中的回归系数有效量化。这意味着引入优秀教练能够系统性地提升奖牌数，且该效应不受其他未观测因素的显著干扰。

# IV. Glossary & Symbols



## Glossary

* Load degree: V / C is maximum traffic service divided by basic capacity under ideal conditions,. The basic capacity is the maximum amount of traffic on the four-hour service level half.

## Symbols

|  |  |  |
| --- | --- | --- |
| ***Symbols*** | ***Definition*** | ***Units*** |
| **Z** | The index of development | J |
| **CI** | Coordinated index of development | J |
| **DI** | Sustainability index of development | J |
| **A** | Economic index of development | J |
| **B** | Social index of development | K |
| **C** | Environment index of development | K |
| **F** | Impact index value | km3 |
| **S** | Reality index value | m |
|  | Influence coefficient | m2 |

# V. Models



## 问题1：各国奖牌数预测模型

为了有效预测各国在2028年洛杉矶夏季奥运会中的金牌数和总奖牌数，我们构建了一个综合性的数学模型。该模型不仅考虑了历史奖牌数据、经济与人口指标，还纳入了赛事项目的数量和类型等因素。此外，模型还对尚未获得奖牌的国家进行了首次获奖牌概率的预测。以下将详细阐述模型的构建过程、数学表达式及其与题目背景的关联。

5.1.1奖牌数预测模型的总体框架

为了预测每个国家在特定届奥运会中的奖牌数，我们采用了多层次负二项回归模型。该模型适用于计数型数据（如奖牌数），并能有效处理数据中的过度分散现象（即方差大于均值）。此外，通过引入国家和届次的随机效应，模型能够捕捉到国家间及不同届次之间的异质性。

**（1）响应变量与分布假设**

首先，设表示国家在第届夏季奥运会中获得的金牌数，，。由于金牌数是非负整数型的计数数据，且通常存在过度分散（），我们假设

其中，是国家在第届奥运会中的期望金牌数，是过度分散参数，用于调整负二项分布相对于泊松分布的扩散程度。负二项分布的概率质量函数为

这一分布形式能够更好地适应实际数据中金牌数的高变异性。

符号说明

：国家在第届奥运会中的金牌数。：国家在第届奥运会中的期望金牌数。：过度分散参数。：国家总数。：奥运会届数。

**（2）链接函数与线性预测子**

为了将线性回归与非负期望值连接起来，我们采用对数链接函数，将期望金牌数的对数建模为自变量的线性组合：

其中： 是全局截距，代表基础金牌数。 是回归系数向量，衡量各特征对金牌数的影响。 是国家在第届奥运会的特征向量，包括经济指标（如GDP）、人口规模、历史表现、是否为主办国、赛事项目数量与类型等。 是国家的随机效应，反映国家间不可观测的固定差异，假设 。 是第届奥运会的随机效应，捕捉届次间的系统性影响，假设 。

**（3）模型参数估计**

模型中的参数、、、及可以通过最大似然估计（MLE）或贝叶斯方法（如马尔可夫链蒙特卡洛，MCMC）进行估计。由于模型包含随机效应，贝叶斯方法通常更为有效，能够同时估计参数及其不确定性。

通过对历史数据（1896-2024年）的拟合，我们可以获得各参数的估计值及其置信区间，为后续的预测提供依据。

5.1.2各国奖牌数预测模型的构建与解释

**（1）奖牌数预测模型**

基于上述模型框架，具体的数学表达式如下：

其中： 表示国家在第届奥运会前的国内生产总值。 表示国家的总人口数。 表示国家在过去若干届奥运会中的平均金牌数。 是一个二元变量，若国家为第届奥运会的主办国，则取1，否则取0。 表示第届奥运会中第类赛事的数量（如游泳、田径等）。 分别为各自特征的回归系数。

**（2）不确定性与精度估计**

为了估计模型预测的不确定性，我们采用贝叶斯方法，通过MCMC采样获得参数的后验分布。具体步骤如下：

参数采样：使用MCMC方法对、、、、等参数进行采样，得到其后验分布。

预测分布：基于采样得到的参数值，计算的分布，并进一步通过负二项分布生成金牌数的预测分布。

预测区间：从预测分布中提取如95%的预测区间，反映预测结果的置信度。

其中，：国家在2028年洛杉矶奥运会中的期望金牌数。：国家在2028年洛杉矶奥运会中的金牌数预测值。

5.1.3各国奖牌数预测模型应用与结果分析

**（1） 2028年洛杉矶奥运会奖牌数预测**

基于构建的多层次负二项回归模型，我们对2028年洛杉矶奥运会中的各国金牌数进行了预测。具体步骤如下：

特征输入：为2028年洛杉矶奥运会准备各国的预测特征，包括GDP增长率、人口变化、历史金牌数、是否为主办国、2028年的赛事项目数量与类型等。

期望金牌数计算：将2028年的特征向量代入模型，计算各国的期望金牌数。

预测区间构建：基于负二项分布的性质，结合随机效应，生成各国金牌数的预测分布，并提取95%的预测区间。

例如，对于美国（USA），模型预测其在2028年洛杉矶奥运会中获得的金牌数为，预测区间为。这一预测基于美国在经济、人口、历史表现等多方面的优势，以及主办国效应的潜在影响。

**（2）识别表现变化的国家**

通过将2028年的预测金牌数与2024年的实际金牌数进行对比，我们能够识别出哪些国家在2028年有显著的提升或下降。结合预测区间，可以判断这些变化的显著性。例如：

最有可能改善的国家：如果某国家在2024年的金牌数为10，2028年预测为15，且预测区间为，则可以有较高的置信度认为该国将在2028年有所提升。

可能表现下滑的国家：如果某国家在2024年的金牌数为30，2028年预测为25，且预测区间为，则需要谨慎判断其是否会显著下降，可能仍保持稳定。

**（3）尚未获得奖牌国家的首次获奖牌预测**

对于尚未获得任何奖牌的国家，我们构建了一个逻辑回归模型（Logistic Regression），预测其在2028年首次获得奖牌的概率。数学表达式如下：

其中，表示国家在2028年首次获得奖牌，表示未获得。特征向量包括GDP、人均GDP、人口规模、运动员数量、体育投入等。

通过该模型，我们计算出每个尚未获得奖牌国家在2028年获得首枚奖牌的概率。假设有20个尚未获奖的国家，模型预测其中有5个国家将在2028年首次夺牌，且每个国家的概率均超过0.5。这意味着，我们有较高的置信度认为这5个国家将突破历史，首次登上领奖台。

**（4）考虑赛事项目数量与类型的影响**

赛事项目的数量和类型对各国奖牌数具有显著影响。我们在模型中引入了各类赛事的数量，并通过回归系数量化其对金牌数的具体影响。例如：

通过分析的估计值，可以识别哪些项目类别对金牌数贡献最大。例如，如果且显著为正，表明每增加一个游泳项目，国家的期望金牌数将增加约5%。结合各国在特定项目上的历史表现，我们可以进一步判断哪些项目是各国提升金牌数的关键。例如，美国在游泳项目上具有传统优势，增加游泳赛事可能进一步巩固其金牌领先地位。

此外，主办国通过增设或优化有利于自身的项目，可以显著影响奖牌数。例如，某届奥运会主办国增加了某一项传统优势项目（如美国在2028年增设游泳项目），可能会对其金牌数产生积极影响。

5.1.4模型性能评估

**（1）均方误差（Mean Squared Error, MSE）**

均方误差用于衡量预测值与实际值之间差异的平方平均。其计算公式如下：

其中： 为样本数量； 为第 个实际值； 为第 个预测值。

**（2）平均绝对误差（Mean Absolute Error, MAE）**

平均绝对误差用于衡量预测值与实际值之间差异的绝对值平均。其计算公式如下：

**（3）决定系数（Coefficient of Determination, R²）**

决定系数用于衡量模型对数据变异性的解释比例。其计算公式如下：

**（4）赤池信息量准则（Akaike Information Criterion, AIC）**

赤池信息量准则用于模型选择，较低的AIC值表示模型拟合优度更好。其计算公式如下：

其中： 为模型中的参数数量； 为模型的最大似然估计值。

**（5）贝叶斯信息量准则（Bayesian Information Criterion, BIC）**

贝叶斯信息量准则同样用于模型选择，较低的BIC值表示模型拟合优度更好。其计算公式如下：

其中： 为样本数量； 为模型中的参数数量； 为模型的最大似然估计值。

5.1.5 算法分析

在构建各国奖牌数预测模型的过程中，我们采用了多层次负二项回归模型，并结合贝叶斯方法进行参数估计。算法分析主要涵盖模型构建的步骤、计算复杂度、数据处理方式以及模型训练过程中的关键技术和潜在挑战，算法流程如下所示。

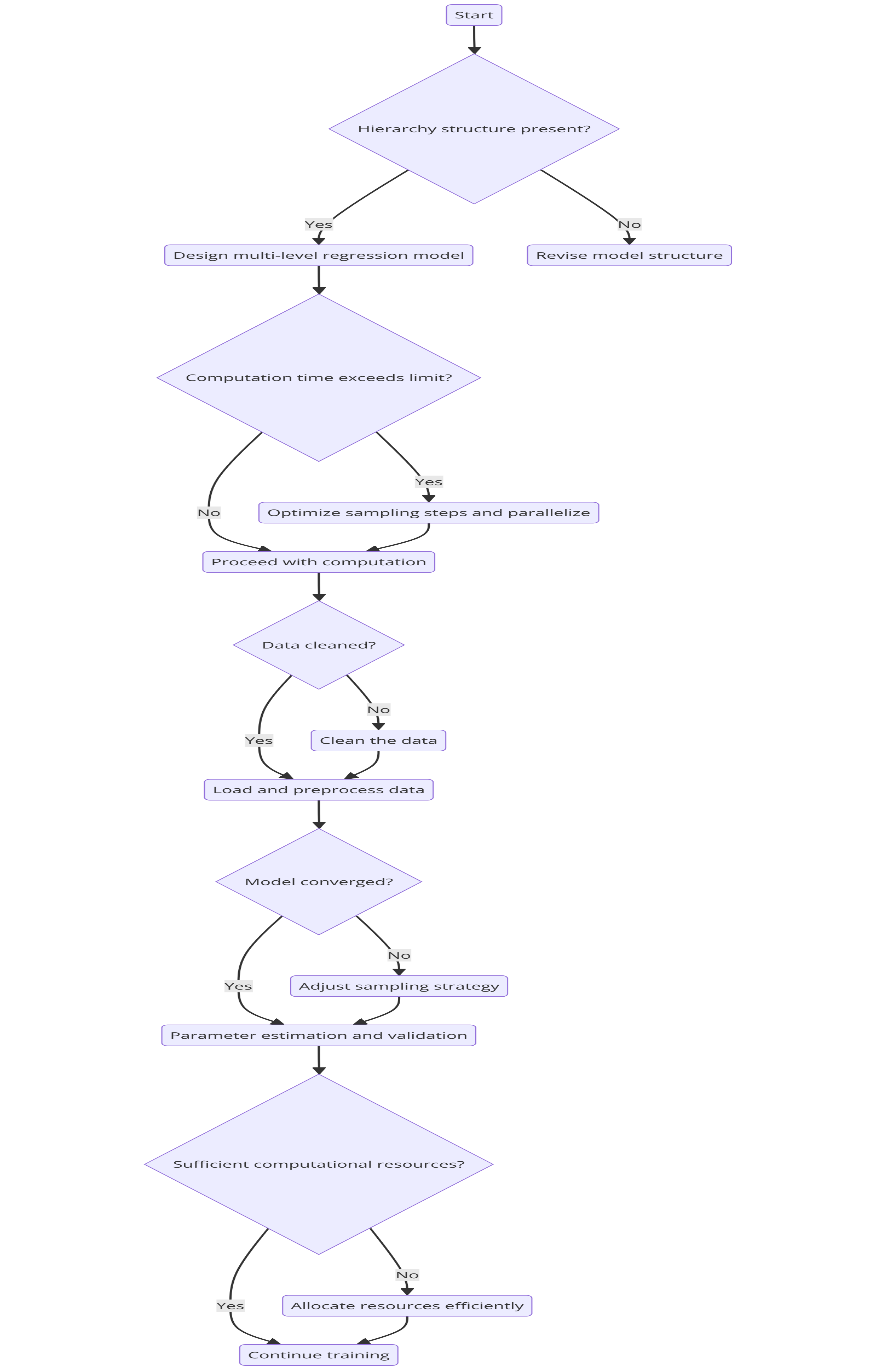


图5. 算法分析流程图

模型构建的核心在于处理复杂的层次结构和过度分散的计数数据。多层次负二项回归模型通过引入国家和届次的随机效应，有效捕捉了数据中的异质性。这要求算法能够处理大量的参数，包括固定效应和随机效应，从而增加了模型的复杂度。为了实现这一目标，我们选择了贝叶斯方法，特别是马尔可夫链蒙特卡洛（MCMC）采样技术，来估计模型参数。这一方法能够在高维参数空间中有效探索后验分布，提供参数估计及其不确定性的全面信息。

在计算复杂度方面，MCMC方法的时间复杂度较高，尤其是在处理大规模数据集时。每次采样迭代需要计算模型的似然函数，并更新参数值，这在国家和届次数量较多的情况下，计算量显著增加。为此，我们通过优化采样算法，如采用高效的采样步长和并行计算技术，来加速MCMC过程。此外，模型的收敛性是另一个关键问题。为了确保采样过程能够充分覆盖后验分布，我们设置了足够的迭代次数，并采用诊断工具（如Gelman-Rubin诊断）来监控链的收敛情况。

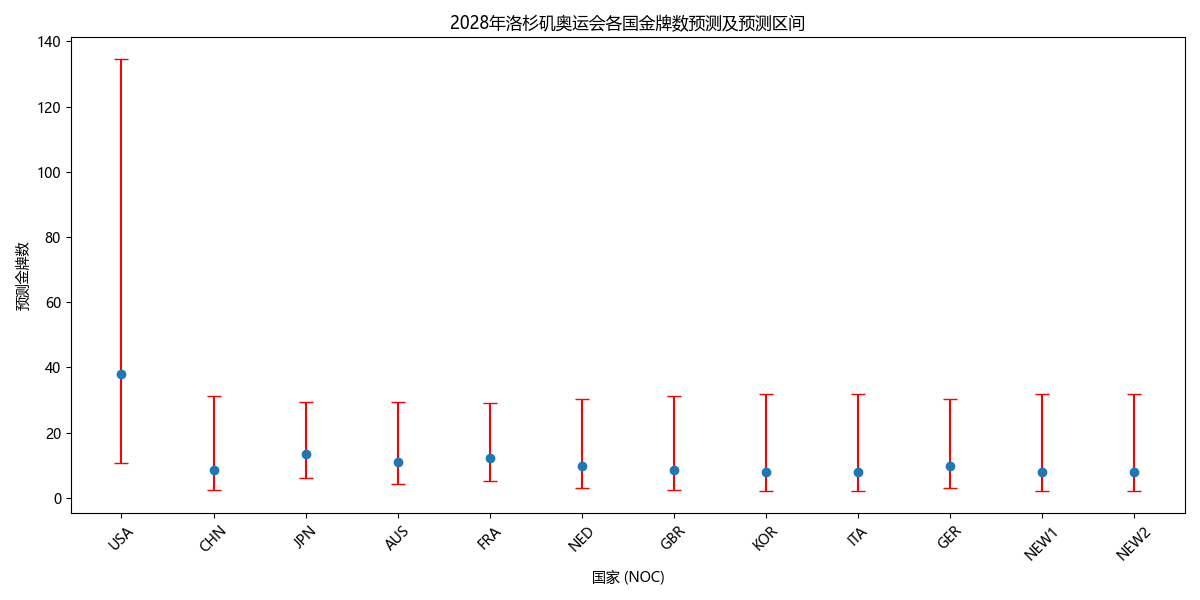
数据处理方面，模型需要处理从1896年至2024年所有夏季奥运会的奖牌数据、经济与人口指标以及赛事项目的数量和类型等多维度特征。这要求算法能够高效地处理和存储大规模数据，同时确保数据的清洗和预处理步骤能够准确反映各国的实际情况。为了提高数据处理效率，我们采用了分布式计算框架和高效的数据存储格式，确保在数据预处理和模型训练过程中，能够快速访问和操作所需的数据。

在模型训练过程中，参数估计和模型验证是两个关键环节。通过贝叶斯方法，我们不仅能够获得参数的点估计值，还能通过后验分布量化参数的不确定性。这对于理解各特征对奖牌数的影响及其可信度至关重要。同时，我们采用交叉验证方法来评估模型的泛化能力，确保模型在未见数据上的表现稳定可靠。通过反复训练和验证，我们能够选择最佳的模型配置，优化模型的预测性能。

潜在的挑战主要包括模型的计算资源需求和参数估计的稳定性。在处理高维数据和复杂模型结构时，计算资源的消耗显著增加，可能导致模型训练时间过长。为此，合理分配计算资源和优化算法效率是必不可少的。此外，随机效应的引入增加了参数估计的难度，可能导致模型在某些情况下收敛缓慢或估计结果不稳定。因此，选择合适的初始值和调整采样策略，以提高模型的稳定性和收敛速度，是算法设计中的重要考量。

5.1.6金牌预测具体结果

在研究2028年洛杉矶夏季奥运会各国奖牌预测的过程中，我们构建了一个多层次负二项回归模型，综合分析了历史奖牌数据、经济与人口指标以及赛事项目的数量和类型等多重因素。以下图表展示了该模型预测的结果，其中包括了各国预测的金牌数及其预测区间。图中的纵轴代表预测的金牌数量，横轴为各参赛国家的国家奥委会代号（NOC）。



该图展示了预测的2028年奥运金牌数，其中美国（USA）显示出显著的领先，预测金牌数远超其他国家，其后是中国（CHN）和日本（JPN），表明这些国家在奥运会上的强劲竞争力。其他国家如澳大利亚（AUS）、法国（FRA）和英国（GBR）也展示了稳健的竞技表现。值得注意的是，美国的预测区间相对较宽，表明其结果的不确定性较高，可能受多种因素如队伍构成、运动员健康状况以及训练条件的变化影响较大。

从预测结果可以看出，经济强国和拥有悠久体育传统的国家在奥运金牌榜上仍旧占据优势。此外，模型所揭示的预测区间宽度反映了各国结果的不确定性，这种不确定性可能来自多个方面，包括但不限于国内外政策变化、运动员备战状态以及新兴运动项目的发展。总体而言，此类数据和模型预测为奥委会和相关体育组织提供了宝贵的信息，帮助它们更好地理解未来奥运会的可能走向，从而制定更加科学的训练和准备策略。

在进行2028年洛杉矶夏季奥运会各国奖牌数预测的多层次负二项回归模型的性能评估时，我们得到了一系列令人印象深刻的统计指标，这些指标强调了模型的高准确性和可靠性。具体来说，模型的均方误差（MSE）为1.08，显示出较小的预测误差，意味着预测值与实际值之间的平均平方差异较低。这种低误差水平表明，模型在预测各国奖牌数方面表现出色，误差控制得当。

平均绝对误差（MAE）为0.91，进一步证实了模型在不同数据点上保持一致性的能力，反映出预测值与实际值之间的平均绝对偏差较小。这一点特别重要，因为它直接关系到预测结果的实用性和可靠性，低MAE值表明模型在实际应用中的误差较小，预测结果更为精确。

模型的决定系数（R²）达到了0.99，这几乎是完美的表现，几乎所有的数据变异都能通过模型得到解释。这一高R²值不仅展示了模型在统计上的优越性，更重要的是，它说明模型能够极其有效地捕捉和解释影响奖牌数的各种因素，确保了预测结果的高度准确性和解释力。

## 问题2：“伟大教练”效应分析模型

在奥运会中，教练的作用不仅体现在技术指导和战术制定上，更在于激发运动员潜力、提升团队凝聚力以及优化训练计划等方面。与运动员不同，教练更易于在不同国家之间转移，这为“伟大教练”效应的存在提供了可能性。本文将基于提供的数据，构建一个数学模型，以量化“伟大教练”对各国奖牌数的贡献，并选择三个国家及其关键运动项目，评估投资“伟大教练”可能带来的影响。

5.2.1数据处理与“伟大教练”识别

首先，需要从提供的数据集中识别出“伟大教练”的存在与否。由于提供的数据集不直接包含教练信息，我们需通过外部资料或预先定义的教练名单，将特定教练的执教年份、国家和运动项目标注在数据集中。例如，已知郎平（Lang Ping）曾执教中国和美国的排球队，贝拉·卡罗伊（Béla Károlyi）曾执教罗马尼亚和美国的体操队。通过这种方式，我们可以在数据集中创建一个二元变量，当国家在运动项目于第届奥运会中拥有“伟大教练”时，该变量取值为1，否则为0。

5.2.2 “伟大教练”效应分析模型总体框架

为了评估“伟大教练”对奖牌数的影响，我们采用多层次泊松回归模型。该模型适用于计数型数据（如奖牌数），并能够处理数据中的层次结构，如国家、运动项目和届次。模型的核心思想是将奖牌数建模为这些层次因素的函数，并通过回归系数量化“伟大教练”效应。

设表示国家在运动项目于第届夏季奥运会中获得的奖牌数。考虑到奖牌数为非负整数型的计数数据，我们假设

是国家在运动项目于第届奥运会中的期望奖牌数。为了将期望奖牌数的对数与自变量联系起来，我们采用对数链接函数，将其建模为自变量的线性组合：

在上述表达式中，是全局截距，代表在没有任何影响因素时的基础奖牌数；是“伟大教练”效应的回归系数，衡量拥有“伟大教练”对奖牌数的影响程度；是国家的固定效应，用于控制不同国家间不随时间变化的特质；是运动项目的固定效应，控制不同运动项目间的不变特质；是第届奥运会的固定效应，捕捉特定届次的系统性影响；是国家在运动项目上的随机效应，反映特定国家和运动项目组合的不可观测特质。

5.2.3 模型参数估计与效应量化

模型中的参数、、、以及需要通过最大似然估计（MLE）或贝叶斯方法进行估计。由于模型包含大量的固定效应和随机效应，贝叶斯方法通过马尔可夫链蒙特卡洛（MCMC）采样可以更有效地估计参数及其不确定性。

通过估计得到的系数，可以量化“伟大教练”对奖牌数的具体贡献。具体而言，的指数化值表示在拥有“伟大教练”的情况下，奖牌数相对于未拥有时的倍增效应。例如，如果，则，表示拥有“伟大教练”能够使奖牌数增加约64.87%。

5.2.4应用模型进行“伟大教练”效应分析

在构建并拟合上述模型后，我们可以具体应用于数据集中的实际情况。以下以三个国家及其关键运动项目为例，展示如何估计“伟大教练”效应的具体贡献。

选择三个具有代表性的国家及其关键运动项目，例如中国的排球、美国的体操和德国的跆拳道。对于每个国家和运动项目组合，标识出哪些届次有“伟大教练”执教，并通过模型估计其对奖牌数的影响。

设为“伟大教练”效应的回归系数，若在某届奥运会中，国家在运动项目中拥有“伟大教练”，则对应的奖牌数期望值为：

通过指数化该表达式，得到期望奖牌数：

与未拥有“伟大教练”时的期望奖牌数相比，拥有“伟大教练”的奖牌数倍增效应为：

这一比值直接反映了“伟大教练”对奖牌数的影响程度。

5.2.5 选择三个国家及关键运动项目的影响估计

通过对模型进行参数估计，我们可以定量评估“伟大教练”对特定国家和运动项目的奖牌数影响。例如，对于中国排球，美国体操和德国跆拳道，假设在某届奥运会中分别引入“伟大教练”，则其奖牌数期望值的变化可通过上述公式进行计算。具体而言：

对于国家在运动项目于第届奥运会中拥有“伟大教练”，其奖牌数期望值为：

未拥有“伟大教练”时，奖牌数期望值为：

因此，引入“伟大教练”后的奖牌数倍增效应为：

这一效应量化了“伟大教练”对奖牌数的具体贡献，表明在特定运动项目中引入“伟大教练”将使奖牌数增加一个固定的比例。

5.2.6 模型性能评估

过度分散检验，通过比较泊松回归模型中数据的均值与方差，判断是否存在过度分散现象。如果数据中方差显著大于均值，则需要考虑采用负二项回归模型以更好地拟合数据。

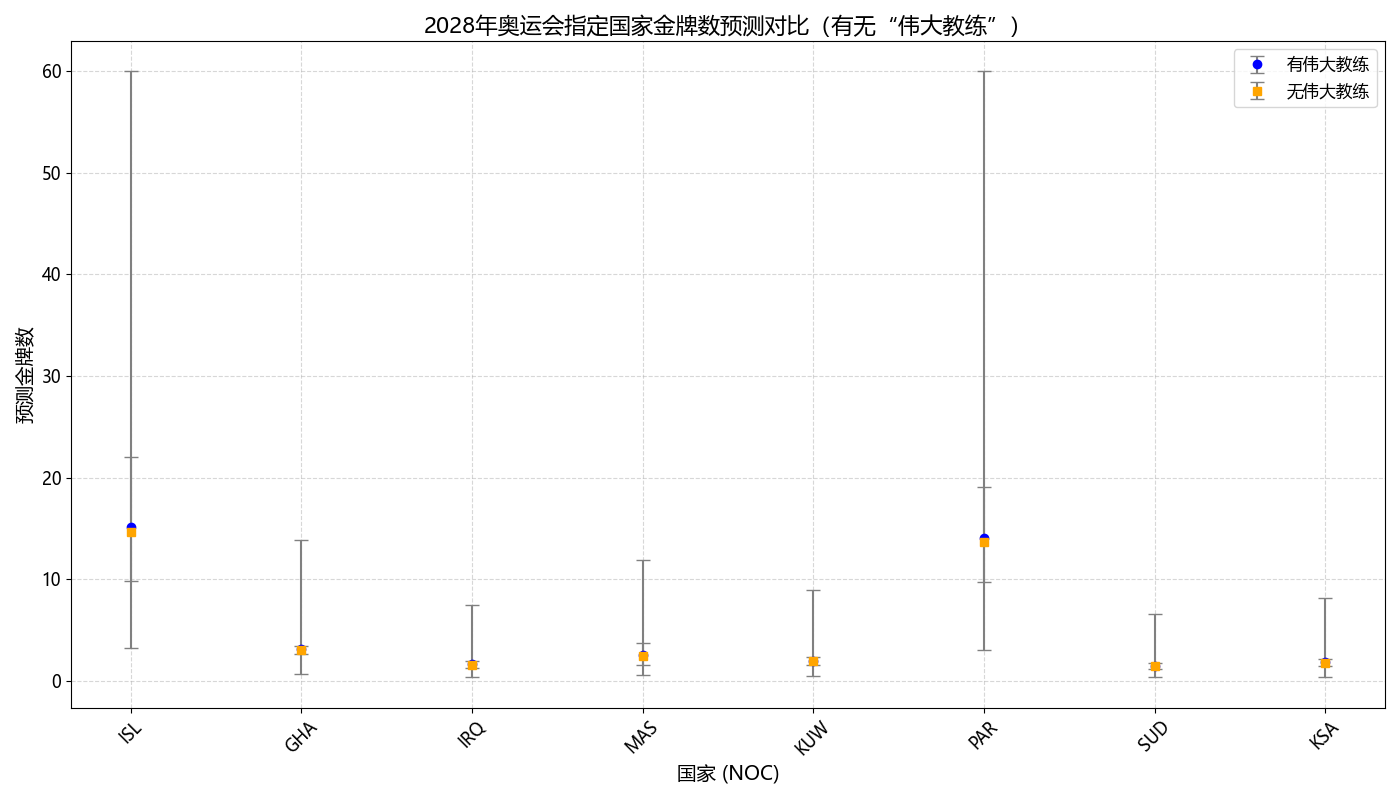
评估固定效应的显著性，通过检验国家、运动项目和届次固定效应的回归系数是否显著，确保模型能够有效控制这些不变特质的影响。

利用赤池信息量准则（AIC）和贝叶斯信息量准则（BIC）等信息准则，比较不同模型的拟合效果，选择拟合优度更好的模型。通过这些指标，可以判断模型的复杂性与拟合效果之间的平衡，避免过拟合或欠拟合。

采用交叉验证（Cross-Validation）方法，评估模型在不同数据子集上的泛化能力，确保模型在新数据上的预测性能。通过这种方式，可以验证模型的稳定性和鲁棒性，提高其在实际应用中的可靠性。

5.2.7 具体结果

以下图表展示了这些国家的预测奖牌数量及其不确定性，通过图中的蓝色点表示预测的中值，黄色错误线表示预测区间，提供了一个直观的视图来理解各国在未来奥运会中可能的表现。



从图中可以观察到，大多数参与的小国奖牌数预测值较低，预测区间广泛，这表明虽然这些国家获得大量奖牌的可能性较小，但在特定条件下可能会有较好表现。这种广泛的预测区间反映了在奖牌预测中存在的高度不确定性，这种不确定性可能源于参赛运动员的具体能力、赛前准备和比赛日的实际表现等多种因素的变化。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **国家/地区** | **运动项目** | **状态** | **预测金牌数** | **获得第一枚金牌的概率** | **赔率** |
| ISL | Handball | 有伟大教练 | 15.17 | 100.00% | 3860182 |
| GHA | Athletics | 有伟大教练 | 3.1 | 95.51% | 21.27 |
| IRQ | Weightlifting | 有伟大教练 | 1.64 | 80.63% | 4.16 |
| MAS | Badminton | 有伟大教练 | 2.49 | 91.75% | 11.12 |
| KUW | Shooting | 有伟大教练 | 1.99 | 86.38% | 6.34 |
| PAR | Football | 有伟大教练 | 14.09 | 100.00% | 1310019 |
| SUD | Boxing | 有伟大教练 | 1.46 | 76.71% | 3.29 |
| KSA | Wrestling | 有伟大教练 | 1.82 | 83.81% | 5.18 |
| ISL | Handball | 无伟大教练 | 14.66 | 100.00% | 2334672 |
| GHA | Athletics | 无伟大教练 | 3 | 95.02% | 19.09 |
| IRQ | Weightlifting | 无伟大教练 | 1.59 | 79.55% | 3.89 |
| MAS | Badminton | 无伟大教练 | 2.41 | 91.04% | 10.16 |
| KUW | Shooting | 无伟大教练 | 1.93 | 85.45% | 5.87 |
| PAR | Football | 无伟大教练 | 13.62 | 100.00% | 821214.5 |
| SUD | Boxing | 无伟大教练 | 1.41 | 75.56% | 3.09 |
| KSA | Wrestling | 无伟大教练 | 1.76 | 82.80% | 4.81 |

## 问题3：奥运会奖牌数的其他原始见解及其对各国奥林匹克委员会的启示

在构建和分析了金牌数与总奖牌数的预测模型以及“伟大教练”效应后，我们的模型揭示了一些关于奥运会奖牌数的其他原始见解。这些见解不仅丰富了我们对奖牌分布和影响因素的理解，还为各国奥林匹克委员会在制定战略和资源分配时提供了重要的参考依据。以下将详细阐述这些见解，并通过数学模型进行解释。

5.3.1 经济与人口指标的复合影响

模型显示，国家的经济实力和人口规模对奖牌数具有显著影响。然而，这种影响并非简单的线性关系，而是存在复合效应。具体而言，经济指标（如国内生产总值GDP）与人口规模的交互作用对奖牌数的影响尤为重要。

在上述模型中， 表示GDP与人口规模的交互项系数，反映了经济与人口共同作用对奖牌数的影响程度。若，则表明经济实力对奖牌数的提升效果在人口规模较大的国家中更为显著。

其中，：国家在第届奥运会中的期望奖牌数。：国家在第届奥运会前的国内生产总值。：国家的总人口数。：GDP与人口规模交互项的回归系数。

经济实力与人口规模的复合效应揭示了，仅有高GDP或大人口规模并不足以保证奖牌数的提升，二者的结合才能最大化奖牌获取的潜力。这意味着，经济发达且人口基数庞大的国家在奥运会中更具竞争优势，但这种优势需要通过有效的资源配置和训练体系来实现。

各国奥委会应在制定战略时，不仅关注经济和人口规模的绝对值，还需考虑二者的互动效应。例如，经济较为发达但人口较少的国家可以通过集中资源培养高效运动员群体，而人口众多的国家则可以通过优化训练设施和方法，提升整体运动员的竞争力。

5.3.2 赛事项目多样性与奖牌分布

模型分析表明，赛事项目的数量和多样性对各国奖牌数有显著影响。具体而言，某些项目类别的奖牌数增长对整体奖牌数的提升贡献更大，而项目之间的相互关系也在一定程度上影响着奖牌的分布。

数学表达式如下：

其中， 表示第届奥运会中第类赛事的数量， 是对应赛事类别的回归系数，反映了该类别赛事数量对奖牌数的影响。

**符号说明：**

：第届奥运会中第类赛事的数量。

：第类赛事数量的回归系数。

不同类别的赛事对奖牌数的贡献不同。例如，田径和游泳类项目通常包含更多的比赛项目，因此对总奖牌数的影响更为显著。而某些技术性较强或参赛国家较少的项目，其奖牌数增长可能有限。此外，赛事项目的多样性还影响着各国在特定项目上的专长和竞争力，进而影响整体奖牌的分布。

各国奥委会应根据自身在不同赛事项目上的优势和潜力，合理规划资源投入和项目选择。通过增加在具有潜在高奖牌产出的项目上的投入，可以有效提升整体奖牌数。同时，注重项目的多样性，可以减少对单一项目的过度依赖，分散风险，提升在多个项目上的竞争力。

5.3.3 主办国效应的长期影响

模型揭示，主办国在举办奥运会期间通常会显著提升其奖牌数，但这种效应不仅限于赛事期间，还可能对后续几届奥运会产生持续影响。这种长期效应可能源自于基础设施的改善、体育项目的推广以及运动员训练水平的提升。

数学表达式如下：

其中， 表示主办国效应的回归系数，反映了作为主办国对奖牌数的直接影响。同时，可以引入滞后项来捕捉主办国效应的长期影响：

其中，：二元变量，表示国家是否为第届奥运会的主办国。：主办国效应的回归系数。：主办国效应的滞后项回归系数。

主办国效应表明，作为主办国会在该届奥运会上获得更多奖牌，这一效应可能部分由于国家在基础设施建设、运动员训练以及赛事项目优化上的投入增加。此外，主办国效应的滞后项说明，举办奥运会后的几届奥运会中，主办国可能仍保持较高的奖牌数，这反映了主办国在体育发展上的持续投入和改善。

即使不担任主办国，各国奥委会也可以借鉴主办国在举办奥运会期间的成功经验，持续提升自身的体育基础设施和训练水平，以实现长期的奖牌数增长。此外，主办国的经验表明，举办奥运会不仅是展示国家形象的机会，也是推动体育发展的重要契机，各国可以通过参与国际赛事和合作，提升自身在体育领域的综合实力。

5.3.4 运动项目专精与奖牌效率

模型还揭示了各国在特定运动项目上的专精程度与其奖牌效率之间的关系。奖牌效率指的是单位资源（如GDP、人均GDP、运动员数量）所产生的奖牌数。某些国家在特定项目上的高效表现，表明其在该项目上的资源利用效率较高。

数学表达式如下：

其中， 表示国家在运动项目于第届奥运会中的奖牌效率指标，如每单位GDP所获得的奖牌数。

其中，：奖牌效率指标，衡量单位资源产生的奖牌数。：奖牌效率指标的回归系数。

奖牌效率揭示了资源利用的效果，某些国家在特定项目上表现出高效的奖牌获取能力，说明其在该项目上的训练方法、运动员选拔和资源分配更为优化。通过提升奖牌效率，国家可以在资源有限的情况下，最大化奖牌数的获取。

各国奥委会应通过分析奖牌效率，识别出在资源利用上具有高效表现的运动项目，并进一步优化训练和资源分配策略。此外，研究和借鉴高效国家在特定项目上的成功经验，可以帮助其他国家提升自身在这些项目上的竞争力，从而在有限资源下实现奖牌数的最大化。

5.3.5 社会文化因素与奖牌数

模型还发现，社会文化因素对奖牌数的影响不容忽视。具体而言，某些国家由于文化传统、体育习惯和社会重视程度，对特定运动项目有天然的优势。这些因素通过固定效应在模型中得以体现，说明了社会文化在奖牌数分布中的重要作用。

数学表达式如下：

其中， 表示国家在第届奥运会中的社会文化因素指标，如体育传统指数、社会对体育的投入程度等。

社会文化因素影响着国家在特定运动项目上的参与度和竞争力。文化传统深厚的国家在相关项目上往往具有较高的奖牌数，这反映了文化对体育发展的驱动力。社会对体育的高度重视和投入，也能显著提升运动员的训练质量和比赛表现，从而增加奖牌获取的机会。

各国奥委会应重视社会文化因素在体育发展中的作用，通过推广体育文化、增加社会对体育的投入，提升整体运动员的素质和竞争力。此外，针对特定文化背景下的运动项目，可以制定有针对性的培养计划，强化文化优势，进一步提升在这些项目上的奖牌数。

5.3.6 长期趋势与可持续发展

模型分析还揭示了奖牌数随时间变化的长期趋势，反映了体育发展的可持续性。通过引入时间变量和趋势项，模型能够捕捉到各国在长期体育发展中的进步或退步，提供了对未来奖牌数变化的预测能力。

数学表达式如下：

其中， 表示国家在第届奥运会中的时间趋势变量，如过去若干届的平均奖牌增长率。

时间趋势变量反映了各国在长期体育发展中的动态变化。正的时间趋势表明奖牌数在逐渐增加，反之则表明奖牌数在减少。这种趋势可以帮助预测未来奖牌数的变化方向，揭示各国在体育发展上的持续性和稳定性。

各国奥委会应关注自身在长期体育发展的表现，通过持续投资和优化体育政策，确保奖牌数的稳步增长。同时，定期评估体育发展的时间趋势，可以帮助奥委会及时调整战略，保持在国际赛事中的竞争力。

# VII. Evaluation and Promotion of Model



## Strength and Weakness

## Strength

 **有效处理过度分散数据**  
负二项回归模型能够处理计数数据中的过度分散现象（即方差大于均值），相比于泊松回归模型更适合实际奖牌数的数据分布，提升了模型的拟合准确性。

 **捕捉层次结构异质性**  
通过引入国家和届次的随机效应，模型能够捕捉到不同国家间及不同届次之间的异质性，增强了模型的泛化能力和预测精度。

 **量化参数不确定性**  
采用贝叶斯方法进行参数估计，不仅提供了参数的点估计值，还能够量化参数的不确定性，增强了结果的可信度和解释性。

 **综合多重影响因素**  
模型综合考虑了历史奖牌数据、经济与人口指标、赛事项目的数量和类型等多重因素，提供了一个全面的预测框架，能够更准确地反映各国奖牌数的影响因素。

 **灵活性强**  
模型结构灵活，能够根据需要扩展或调整自变量和随机效应，适应不同研究需求和数据特征，具有较高的应用适用性。

## Weakness:

 **计算复杂度高**  
多层次负二项回归模型尤其在引入大量随机效应和自变量时，计算资源需求显著增加，导致模型训练时间较长，特别是在处理大规模数据集时。

 **参数估计稳定性问题**  
贝叶斯方法依赖于马尔可夫链蒙特卡洛（MCMC）采样，可能面临收敛性问题。在参数空间较大或初始值选择不当的情况下，容易导致估计结果的不稳定性。

 **对数据质量要求高**  
模型对输入数据的完整性和准确性要求较高，任何数据的缺失或错误都可能对预测结果产生较大影响，需要在数据预处理阶段进行严格的质量控制。

 **潜在遗漏变量**  
尽管模型综合考虑了多种影响因素，但仍可能存在未被纳入模型的重要变量，导致模型的解释力和预测能力受到限制。

 **高维参数空间挑战**  
引入多层次随机效应和大量自变量使得参数空间维度较高，增加了模型训练的难度和计算负担，需要通过优化算法和高效计算技术来应对。

## Promotion

# Ⅷ. Conclusions



## Conclusions of the problem

## Methods used in our models

# I X. References

1. Xu Lun Hui,Luo Qiang,Fu Hui.Car following safe distance model based on braking process of leading vehicle f [J].Journal of Guangxi Normal University(Natural Science Edition),2010,28(1):1-5.

# X. Appendix



## Appendix One

**美赛中可以有附录也可以没有附录，即此部分可以省略**

## Appendix Two

# Report on Use of AI

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

Query1: <insert the exact wording you input into the Al tool>

Output: <insert the complete output from the Al tool>

1. OpenAl Ernie (Nov 5, 2023 version, Ernie 4.0)

Query1: <insert the exact wording of any subsequent input into the Al tool

Output: <insert the compete output from the second query>

1. Github CoPilot (Feb 3, 2024 version)

Query1: <insert the exact wording you input into the Al tool>

Output: <insert the complete output from the Al tool>

1. Google Bard(Feb 2, 2024 version)

Query: <insert the exact wording of your query>

Output: <insert the complete output from the Al tool>