Mathematical Modeling and Algorithm Application
ISSN: 3006-0842 | Vol. 5, No. 2, 2025

A Study on the Factors Affecting the Number of Olympic Medals Based on a Multivariate Negative Binomial Model

Yanhui Li[†, *], Xinyao Liu[†], Ke Chen[†]

Xiangtan University, Xiangtan, China

[*] **Corresponding author**: Yanhui Li (Email: 3491376880@qq.com)

[†]**These authors also contributed equally to this work.**

Abstract: This paper focuses on a systematic study of the factors influencing the number of medals won by countries at the 2028 Olympics. By constructing a multivariable negative binomial mixed-effects model and combining it with hierarchical regression analysis, the study delves into the mechanisms through which multidimensional factors influence medal counts. The research introduces interaction variables between countries and sports, quantitatively revealing each country's core strengths in specific sports. Additionally, it finds that negative regression coefficients for certain sports can indicate a country's weaker areas in sports. When examining the 'great coach' factor, specific cases confirm that coaches have differing impacts on various events, with more significant returns on investment in events with multiple medals. Additionally, the study quantifies the influence of socio-economic and geographical factors, clarifying that sports culture and historical background have the greatest impact, while population size has a relatively minor influence. This research provides a multi-dimensional theoretical basis for understanding the distribution mechanism of Olympic medals and formulating sports development strategies.

Keywords: Multi-Dimensional Negative Binomial Mixed-Effects Model, Hierarchical Regression Method, Interaction Variable, Olympic Medals.

1. Introduction

As the world's premier sporting event, the number of medals won by countries at the Olympic Games has long been a key topic of sports research. Current academic studies on the factors influencing Olympic medals have covered both macro-level dimensions such as national economy and population structure, as well as micro-level aspects like event characteristics. However, there remain significant gaps in integrated analyses of the interactive effects of multiple factors and in quantitative research on implicit factors such as sports culture [1-2].

This study takes the 2028 Olympics as its research object, based on historical medal data and a multi-dimensional analysis framework. By constructing a multi-dimensional negative binomial mixed-effects model and combining it with hierarchical regression methods, the study systematically explores the mechanisms through which multi-dimensional factors such as countries, events, athletes, and host nations influence medal counts [3-4]. The study particularly focuses on the impact of the interaction between countries and events on the formation of dominant events, revealing the core competitive events and sports weaknesses of different countries by introducing interaction variables [5]. It also analyses the differentiated impact of 'great coaches' on different events through case studies, verifying the varying benefits of coach resource investments in multi-medal events [6]. Additionally, it quantitatively analyses the contribution of socio-economic and geographical factors, clarifying the key role of soft factors such as sports cultural historical backgrounds. This study aims to enhance the theoretical understanding of Olympic medal distribution, providing scientific basis for countries to optimise sports resource allocation and formulate targeted development strategies, thereby enhancing their comprehensive competitiveness in international sports competitions.

2. Prediction based on the multivariate negative binomial model

This paper employs hierarchical regression and the negative binomial model to predict the number of medals each country will win at the 2028 Olympics. Hierarchical regression constructs different models by progressively introducing independent variables, compares model metrics to explore the impact of four factors—country, event, athlete, and host country—on the number of medals, and reveals the relationship between each level and the number of medals to achieve accurate predictions. The negative binomial regression, as a generalized linear model for handling count data, is suitable for data such as medal counts, which are non-negative integers with excessive dispersion (variance greater than the mean). By introducing new parameters to control the variance-to-mean ratio and combining the advantages of hierarchical regression, a multivariate negative binomial mixed-effects model can be constructed to effectively handle large-scale excessively dispersed data and be used for medal count prediction.

2.1. Model construction

In order to accurately predict the medal rankings for the 2028 Olympic Games, given the excessive dispersion of medal data and based on multi-dimensional big data considerations such as historical Olympic medal counts, the number of participating athletes, and the number and types of events, this paper constructs a multivariate negative binomial mixed effects model. And the fixed effects include total events, total participants, and host country, the random effects

include country, discipline, and country-discipline. In addition, considering that directly predicting the total number of Olympic medals for each country may result in significant errors, this paper adopts a project-specific prediction strategy to improve the accuracy and interpretability of the model: first, the number of medals for each country in each individual project is predicted, and then the predicted values for each project are accumulated to obtain the total number of medals for each country.

$$M_{i,j,k} = \left(\text{Gold}_{i,j,k}, \text{Silver}_{i,j,k}, \text{Bronze}_{i,j,k} \right)^T \quad (1)$$

$$\log\left(m_{i,j,k}\right) = \mathbf{x}_{i,j,k}^T \beta + u_i + v_{j,k} + w_k + k + \delta_j \quad (2)$$

Where $m_{i,j,k}$ represents the number of each type of medal (gold, silver, bronze) in the i country, the j year, and discipline k. $X_{i,j,k}$ represents the covariates in the i country, the j year, and the k event, including: total participants represent the total number of athletes in the i country, the j year, and the k event. And β represents the regression coefficient, u_i represents the random effect at the country level, w_k represents the random effect at the discipline level, δ_j represents host country effect in year j.

2.2. Model results

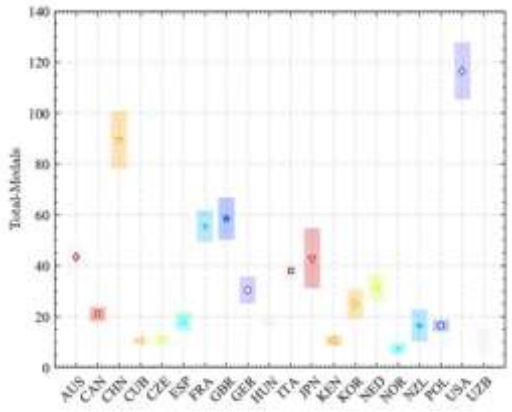

Figure 1. Predicted total number of medals

Figure 1 shows the predicted medal range for the 2028 Olympics. The USA is predicted to win between 110 and 130 medals, making it the only country to break the 100-medal mark, demonstrating its exceptional Olympic competitiveness and continuing its advantage as a sports powerhouse. CHN is predicted to win between 80 and 100 medals, maintaining a high gold medal conversion rate in traditional strongholds such as table tennis, diving, weightlifting, while also improving its competitiveness in athletics and swimming. GBR and FRA are in the 50–70 medal range, leading the second tier. JPN is in the 40–60 medal range, maintaining its lead in the third tier.

3. Exploring the relationship between different factors based on interaction effects

In Section 2, this paper considered the impact of four fixed variables-country, event, athlete, and host country-on the number of medals won by a country. In this section, this paper will focus on the relationship between the country and the number of medals won in various events. By introducing an interaction variable between the two, this paper aims to quantitatively measure the most important events for different countries through regression coefficients. The formula is as follows:

$$\log\left(m_{i,j,k}\right) = \mathbf{x}_{i,j,k}^T \beta + u_i + v_{j,k} + w_k + \lambda_{i,k} + \delta_j \quad (3)$$

Where $\lambda_{i,k}$ represents the country-discipline interaction effects.

3.1. Results analysis

This paper obtained the regression coefficients for each event for all countries. The 10 most important events in each country/region are listed in the Table.1 below.

Table 1. Important discipline

NOC	Important Discipline
CHN	Diving
USA	Swimming
JPN	Judo
FRA	Fencing
GBR	Cycling
AUS	Athletics
NED	Athletics
ITA	Cycling
KOR	Shooting

Taking China as an example, this paper visualized the correlation between the number of medals won in each event and the number of gold medals. The results are shown in the Figure 2.

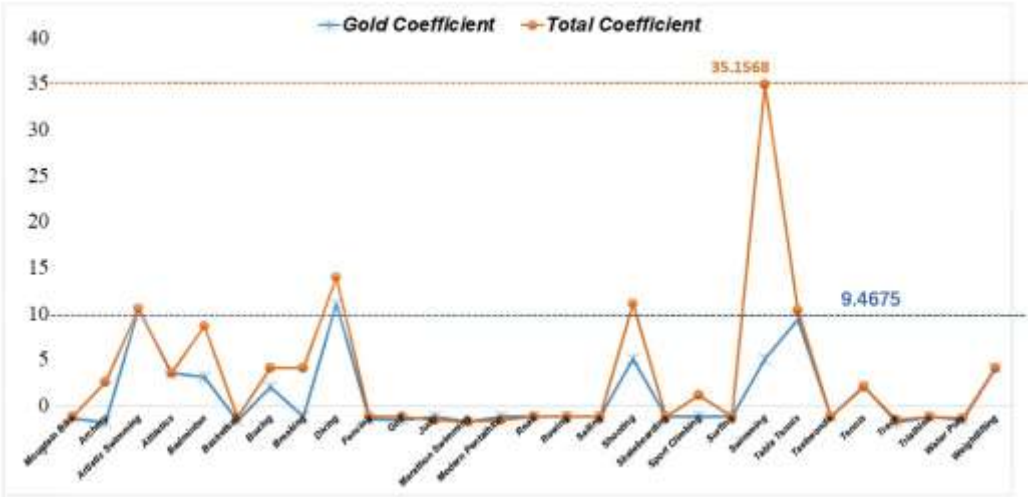

Figure 2. Results in China

For the total number of medals, the swimming event has the highest correlation coefficient, reaching 35.1568. Therefore, this paper can conclude that swimming contributes the most to China's medal count and is the most important event. For the number of gold medals, tennis has the highest correlation coefficient, reaching 9.4675. Thus, it can be inferred that to achieve better results in the Olympics, the country's Olympic committee could consider placing more emphasis on swimming and tennis.

In addition to the highest correlation coefficients, this paper also observes that some events have negative regression coefficients. This paper believes these events may represent the weaker areas for the country. Therefore, the country's Olympic committee should closely monitor the training of these weaker events and increase investment in funding, talent, and resources to improve the overall strength of the country's sports.

4. Analysis of the impact of "Great Coaches"

In Section 2, this paper used a hierarchical analysis and negative binomial regression model to predict the medal counts for each country in the 2028 Olympics. In this, this paper treated country, event, and athlete as fixed variables in the model. By gradually introducing these variables into the model, this paper obtained the contributions of three different variables to the country's medal count.

In this section, the impact of the "Great Coach" factor on the number of medals essentially continues to study the contribution of independent variables to the dependent variable. Therefore, this paper inherited the basic framework of the multivariate negative binomial model and, based on this, introduce "Great Coach" as a fixed variable to analyze its relationship with the number of medals won.

4.1. Data collection

To explore the impact of the "Great Coach" factor on the number of medals for each country, this paper needs to search for possible effective data.

There are two case studies: Lang Ping and Béla Károlyi. Based on these two cases, this paper conducted extensive data queries and collected and summarized the information. The sample data is shown in Figures 3 and Figure 4:

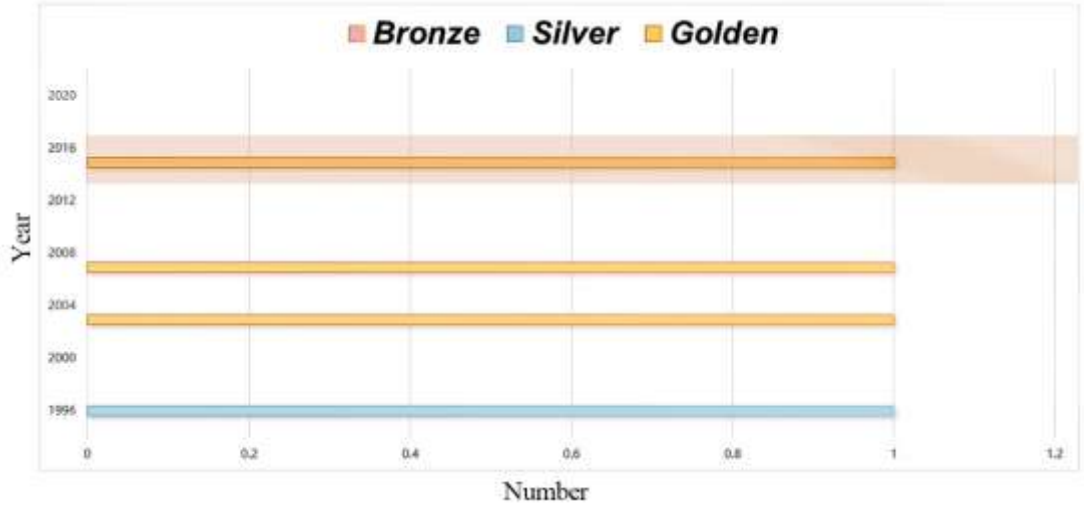

Figure 3. Volleyball in China

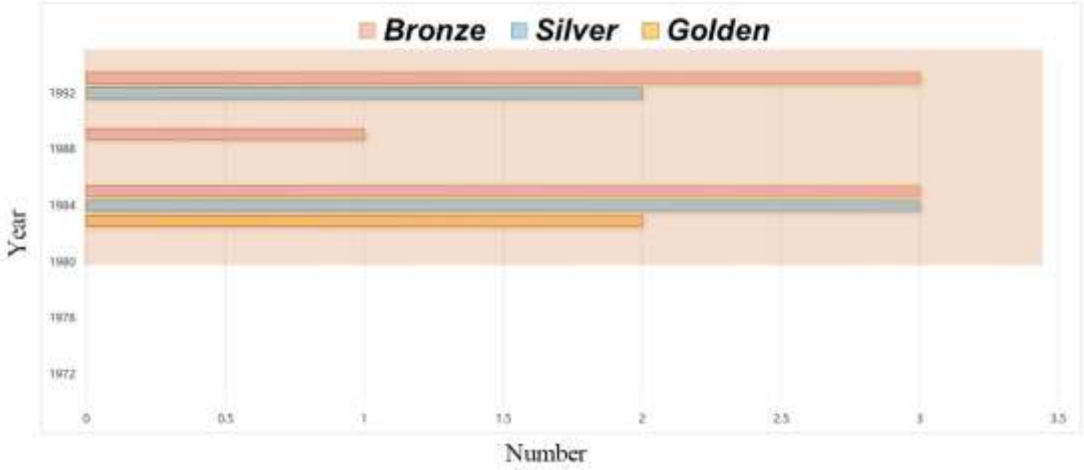

Figure 4. Gymnastic in the US

4.2. Introduction of the regression coefficients for "Great Coaches"

The formula is as follows:

$$\log\left(m_{i,j,k}\right)=\mathbf{x}_{i,j,k}^{T}\beta+u_i+v_{j,k}+w_k+\lambda_{i,k}+\delta_j+C_{i,j} \tag{4}$$

Where $C_{i,j}$ represents "Great Coaches" effect.

4.3. Estimating the impact of "Great Coaches"

Through the calculations of regression coefficients and growth rates in previous section, this paper obtained the following conclusions:

From the regression coefficient table, it can be seen that the regression coefficient for volleyball is 0.1138, while the regression coefficient for gymnastics is 0.1814. This result indicates that the effect of the "great coach" factor on gymnastics is greater than that on volleyball. The reason for this outcome could be that gymnastics, being an individual sport, tends to have a higher number of medals, while volleyball, being a team sport, has fewer medals. Therefore, the "great coach effect" is more prominent in gymnastics than in volleyball.

4.4. Case study

(1) Feature construction

This paper selected China, Russia, and the United States as examples. Here, this paper considered the following factors to assess the effect of the introduction of great coaches:

a. Historical performance: Identify which events may have greater potential or room for improvement.

b. Strong events: Assess the strong events of these countries (e.g., swimming in the United States, table tennis in China) and determine whether these strong events have already reached their peak or if there is room for improvement.

c. Coaching effect: Using historical data, analyze whether great coaches have had a positive impact on specific events in these countries, and whether the medal count in these events can continue to improve.

(2) Impact evaluation

Based on the multivariate negative binomial mixed effects model established in previous section, this paper evaluated the potential contribution of great coaches to the medal count of these countries. The specific approach could be:

a. Estimate the regression coefficients for great coaches.

b. Through model analysis, identify which events will show significant improvement under the guidance of a coach.

c. Compare the medal predictions before and after the introduction of great coaches, and see in which events more medals can be won.

(3) Result analysis

Through the steps above, this paper found that for China, the United States, and Russia, the highest returns from investing in great coaches are achieved in the following events: table tennis, swimming, and judo. The reason for this could be that these events include both individual and team sports, with a higher number of medals available. As a result, the returns from investment are higher due to the greater number of medal opportunities.

5. Exploring the impact of other factors on medal counts

In Section 4, this paper measured the contribution of the "great coach" factor to the number of medals by solving for the coefficients of fixed effects.

In this section, this paper continued with this approach to explore the impact of other potential factors on the medal count of countries.

5.1. Indicator determination

Through literature review and data collection, this paper found that socio-economic factors have a highly significant nonlinear relationship with the number of medals won by countries. In addition, natural geographical factors are also closely related to medal counts. Based on the above analysis, this paper primarily selected the following variables: national per capita GDP, population size, geographical environment, sports culture, and historical background as fixed variables, which were gradually introduced into the model.

(1) Per capita GDP

Per capita GDP represents the overall economic development level of a country and reflects the living standards of its citizens to some extent. When the standard of living is high, the demand for leisure and entertainment also increases, which raises the probability of developing athletic talents. At the same time, per capita GDP indicates a country's economic level-higher economic development typically leads to more investment in sports, thereby attracting more athletic talent.

(2) Population size

In probability theory, if the probability of an event occurring remains constant, the larger the sample space, the more often the event is likely to occur. Based on this logic, this study treats a country's population size as the sample space for medal-winning events—as the population base increases, the potential pool of outstanding athletes also expands, thereby increasing the likelihood of winning medals. Therefore, the impact of population size on the number of Olympic medals has significant statistical significance, and it is thus included as a core variable in the model construction.

(3) Geographical environment

The geographical environment affects the training locations of athletes. When a sport aligns well with a country's geographical conditions, it is more likely that top talent in that sport will emerge, leading to a higher probability of winning medals.

(4) Sports culture and historical background

Culture and history shape the mindset of a country's people. If a country has a rich history of sports culture, its investment in sports at the national level is likely to be greater. On the individual level, people are also more willing to participate in sports, which increases the likelihood of nurturing athletic talent. Therefore, sports culture and history are important factors influencing medal counts.

5.2. Introduction of fixed variables

The formula is as follows:

$$\log\left(m_{i,j,k}\right)=\mathbf{x}_{i,j,k}^{T}\beta+u_i+v_{j,k}+w_k+\lambda_{i,k}+\delta_j+C_{i,j}+G_{i,j}+P_{i,j}+E_{i,k}+T_{i,k} \tag{5}$$

Where $G_{i,j}$ represents the GDP of country i in year j, $P_{i,j}$ represents the population of country i in year j, $E_{i,k}$ represents the environment of country i which is similar to discipline k and $T_{i,k}$ represents the culture of country i to discipline k.

5.3. Results analysis

Based on the historical Olympic medal counts of each country, this paper used the method in previous text to introduce the regression coefficients and solve for them. This paper obtained the extent of the impact of four indicators-national per capita GDP, population size, sports culture, and historical background-on the number of medals won by countries. The results are shown in the Table.2 below:

Table 2. Impacts of different indicators

Indicator	Degree of Impact
GDP	0.00003
Population	0.00000005
Geography	0.07321
Culture	0.15342

From the table, it can be seen that the cultural-related coefficient is the highest at 0.15342, while population size contributes the least to the medal count, with a coefficient of 0.00000005. Therefore, this paper can conclude that in order to increase a country's probability of winning medals, fostering a nationwide culture of sports participation and exercise is a more effective approach.

6. Conclusions

This study systematically analysed the factors influencing the number of medals won by countries at the 2028 Olympics by constructing a multivariable negative binomial mixed-effects model. The study introduced interaction variables between countries and events to quantitatively reveal each country's core strengths, while also finding that negative regression coefficients for certain events could reflect a country's weaknesses in sports. When discussing the 'great coach' factor, case studies showed that coaches had significantly different impacts on different events, and that investing coaching resources in events with multiple medals yielded more significant returns. Additionally, the study quantified the influence of socio-economic and geographical factors, clarifying that sports culture and historical background play a dominant role in medal counts, while factors such as population size have relatively limited effects.

This study not only reveals the mechanisms influencing Olympic medal counts from multiple dimensions and refines the relevant theoretical framework but also provides practical insights for countries in formulating sports development strategies: on the one hand, by identifying core strengths and weaknesses, resource allocation can be optimised; on the other hand, emphasis should be placed on cultivating sports culture and building high-level coaching teams, particularly by strengthening coaching resource investments in multi-medal events to enhance a country's overall competitiveness in international sports. Future research could further expand the dimensions of influencing factors and deepen the exploration of the development mechanisms of emerging sports projects, providing more comprehensive theoretical support for Olympic medal predictions and sports strategic planning.

References

[1] Luan Chuankai. Research on the Evolution of the Global Track and Field Landscape at the Tokyo Olympics and Development Strategies for Chinese Track and Field [J]. Shandong Sports Science and Technology, 2023, 45 (02): 47-52.

[2] Aisike Baituo, Jing Longjun. Analysis of the Distribution Characteristics of Chinese Medals at the Tokyo Olympics [J]. Fujian Sports Science and Technology, 2022, 41 (04): 46-50.

[3] Li Jingcheng, Cai Weibin. Factors Influencing the Usefulness of Online Reviews and Configuration Analysis: Based on Negative Binomial Regression and fsQCA Methods [J]. Resource Development and Market, 2023, 39 (01): 1-8.

[4] Zhuang Yan, Dong Chunjiao, Mi Xueyu, et al. Identification of Accident Black Spots Based on Improved Network Kernel Density and Negative Binomial Regression [J]. Journal of South China University of Technology (Natural Science Edition), 2024, 52 (01): 119-126.

[5] Liang Kang, Shu Yankai. Analysis of Medal Distribution Characteristics in Ice Sports at the Beijing Winter Olympics [J]. Ice and Snow Sports, 2023, 45 (02): 7-10+59.

[6] Li Jilin, Ren Yuanbo, Wang Yang. The Pathways and Mechanisms of Perceptual Coaching Support on the Athletic Performance of Professional Athletes [J]. Journal of Xichang University (Natural Science Edition), 2025, 39 (01): 93-101.