

# 奥运会奖牌榜预测建模实证分析

**摘要** 本文以 1-31 届夏季奥运会的数据为样本, 确定了 17 个国家作为研究对象。首先根据每个国家历届奥运会历史成绩的内在关联性, 利用时间序列预测和灰色预测方法建立了关于奥运奖牌数目的预测模型, 并对这些国家未来一届的奖牌数目作出预测。然后进一步考虑国家综合实力对奥运会成绩的影响, 选取 5 个衡量国家综合实力的指标, 利用非线性回归分析的方法建立了关于国家综合实力和历史成绩的多元非线性回归方程来预测未来成绩。最后, 利用非线性带虚拟变量的多元回归方法进一步分析东道主效应、人种以及国家体制等因素对奥运会成绩的影响, 并对下一届奥运会奖牌前十名的国家给出了预测结果。

**关键词:** 奥运会 奖牌榜 灰色预测 时间序列 回归分析

## 目录

一、引言.....	1
二、数据介绍.....	1
2.1 问题分析.....	1
2.2 模型假设.....	2
2.3 符号说明.....	2
2.4 数据预处理.....	2
三、利用时间序列和灰色预测模型对奥运会奖牌数进行预测.....	3
四、多元回归预测模型的建立.....	8
4.1 影响因素的确立.....	8
4.2 数据获取.....	8
4.3 模型建立.....	8
4.4 模型求解.....	10
4.5 模型检验.....	10
五、模糊综合评价模型对国家进行分类.....	11
5.2 指标等级的分类.....	12
5.3 模糊综合评价模型的建立.....	12
5.4 用模糊综合评价模型求解.....	13
六、模型的评价和推广.....	14
6.1 模型评价.....	14
6.2 模型推广.....	15
总结.....	16
参考文献.....	17
附录.....	18

## 一、引言

奥林匹克运动会发源于两千多年前的古希腊，现代奥运会，是国际奥林匹克委员会主办的世界规模最大的综合性运动会，每四年一届，会期不超过 16 日，是目前世界上影响力最大的体育盛会。自 1896 年举办首届奥运会以来，已举办了 31 届。其中第 6 届因第一次世界大战停办，第 12、13 届因第二次世界大战停办。奥运会的每一项竞技都汇聚了全世界最优秀的运动员，所以它的竞技名词就是人类体能与竞技力的一次记录，奥运会成绩也是一个名族、一个国家意志力的表现，因此近年来对奥运会奖牌榜的预测也吸引了各界学者。吴殿廷（2008）分析了东道主效应并利用灰色预测模型 GM（1,1）预测了 2008 年奥运会北京和美国的金牌数量；王国凡（2010）利用模糊聚类法预测了 2008 年参赛国家的实力，并对参赛国给出了合理的分类；秦瑞瑞（2011）基于 PSO 算法对奥运会奖牌做出了预测分析。David Forrest（2010）通过建立关于金牌数的多元回归模型预测了 2012 年伦敦奥运会的金牌数目。Bernard（2001）分析了参赛国家的人口数与奥运会成绩的关系。李力研（2001）分析了人类种族的差异对体育竞技成绩的影响。本文基于以上研究方法和内容，一方面考虑每个国家历届获得奖牌数之间的内在关联关系，通过建立时间序列预测模型预测了下一届奥运会参赛国家奖牌的数目。另一方面，通过考察诸多其它因素对各参赛国所获奖牌数目的影响，通过建立多元回归方程对下一届奥运会奖牌数目前十名的国家做出预测。

## 二、数据介绍

本文采用的世界各国在历届奥运会上获得的奖牌数来源于国际奥委会官方网站：[www.olympic.org](http://www.olympic.org)，本文首先搜集了 1--31 届奥运会奖牌数排名前 10 的国家（见附录 1），然后综合考虑有些国家的历史变化情况（例如一些国家的解体）以及一些国家相关数据的缺失（古巴的经济发展数据），最终确定了以下 16 个国家作为研究对象：美国、中国、英国、日本、法国、德国、匈牙利、澳大利亚、意大利、韩国、俄罗斯、加拿大、瑞典、芬兰、荷兰、巴西。反映各个国家综合实力的指标，比如人均 GDP（按 2010 年不变价美元）、人口数（百万）、出生时的预期寿命、R&D 研究人员（每百万人）、商品贸易占 GDP 的百分比等数据均来源于世界银行数据库网站：<http://econ.worldbank.org>，本文所用到的这些数据均为举办历届奥运会年底的数据。（见附录二）

### 2.1 问题分析

对于年奥运会奖牌榜的排名情况这个不确定性过程，单纯地对每个国家的历史排名进行线性或非线性拟合在进行预测并不可取。因为得出的每个国家的排名可能相等，对预测排名相等的国家进行排序又将成为一个棘手的问题，另外这种方法没有考虑到竞技体育中各个国家的相互影响。相反的马尔科夫链模型对这类问题具有很好地适应性，它的短板就是需要保证下一届的排名情况只与前一届的排名有关。因此，我们需要先对近几届的奥运会奖牌榜排名情况进行相关性分析，若符合上述条件则按马尔科夫链预测模型的方法可以很好滴解决这个问题；若不满足我们需要马尔科夫链预测模型进行改进，比如分别考虑前几届奖牌榜排名对下一届的影响。

接下来用以往的奥运会奖牌榜的排名数据产生概率转移矩阵，使用马尔科夫链或者改进的马尔科夫链

进行预测，针对预测的偏差，可以用多元线性回归模型进行修正，将其他影响奖牌榜排名的因素纳入考虑范围，建立模型后对已知奖牌榜排名进行检验。对于各国体育水平的分类可以使用模糊评价法，模糊评价的指标为上述奖牌榜预测的排名和其他与体育水平有关的因素，其中各项指标的权重运用熵值法进行确定。

## 2.2 模型假设

结合实际，我们提出了以下假设：

- (1) 两届奥运会排名相关性很小的视为无影响
- (2) 世界各国的人口、平均年龄、GDP 等指标可以代表其体育发展水平，并且收集到的数据均真实，没有人为臆造的情况
- (3) 所选取的指标数据短时间内不会发生变化
- (4) 不考虑体育大国由于各种原因拒绝参加奥运会
- (5) 奖牌榜前十名中的国家不会发生剧变
- (6) 各个国家的运动员都诚信参赛，没有使用药物（如兴奋剂）的作弊情况
- (7) 每个国家的人口、平均年龄、GDP 等指标的值都不会无限增大，都会达到一个最大值

## 2.3 符号说明

符号	含义	符号	含义
$b$	一个国家奖牌榜上的名次	HDT	人文发展指数
$P$	状态转移概率矩阵	$R_i$	综合测度向量
PEO	人口总数	$\alpha$	主观确定的权重向量
MAN	入境旅游人数	$\beta$	综合权重向量
HOST	是否为东道主国家	$e_j$	第 $j$ 个指标的熵值

## 2.4 数据预处理

依照国际奥林匹克委员会的官方奥运奖牌排名榜排列，首先衡量金牌枚数的多少，银牌次之，最后对比铜牌枚数。若两支或以上的队伍在同一时间拥有相同的金、银、铜牌枚数，则排同一名次，而采用国际奥委会发布的国家编码，依国家名字首字母顺序的前后进行排位，我们从新浪网竞技风暴板块得到过去 5 届奥运会奖牌榜排名情况（表 2-1）。

表 2-1 26-30 届奥运会奖牌榜

国家	排名 2012	排名 2008	排名 2004	排名 2000	排名 1996
美国	1	2	1	1	1
中国	2	1	2	3	4
英国	3	4	10	10	36
俄罗斯	4	3	3	2	2

韩国	5	7	9	11	10
德国	6	5	6	5	3
法国	7	10	7	6	5
意大利	8	9	8	7	6
匈牙利	9	21	13	12	12
澳大利亚	10	6	4	4	7
日本	11	8	5	15	23
哈萨克斯	12	29	40	22	24
荷兰	13	12	17	8	15
乌克兰	14	11	12	21	9
古巴	15	28	11	9	8

奖牌榜数据表明相邻两届的奥运会奖牌榜没有太大变化，以 08 年和 12 年来看，奖牌榜前十名中只有日本被排挤到了第 11 位，而匈牙利则后来居上跃居第 9 位。对于 08 年北京奥运会和 04 雅典奥运会来说奖牌榜中的前十名国家没有任何变化只是排名有所改变。年表中数据还表明，届次相隔越多奖牌榜排名变化越大，也就是说与该次奥运会相隔时间越长的奥运会排行榜与这次奥运会排行榜的关联越小。因此，我们计算出两届奥运会排行榜间的相关系数（表 2-2）以量化这种关系。

表 2-2 过去 5 届奥运会奖牌榜排名相关系数

相关系数	2012	2008	2004	2000	1996
2,012	1				
2,008	0.726	1			
2,004	0.513	0.762	1		
2,000	0.627	0.603	0.728	1	
1,996	0.189	0.253	0.501	0.578	1

表中结果反映出的规律与预期基本相符，对于相差 1 届的两届奥运会，奖牌榜排名间的相关系数为 0.7 左右；时间间隔为 8 年的两届奥运会，奖牌榜排名间的相关系数为 0.5 左右；相差 2 届的两届奥运会，奖牌榜排名间的相关系数为 0.3 左右；而相差 3 届的两届奥运会奖牌榜排名间的相关系数仅为 0.189，相关性极小可忽略不计。综上可得出结论，一届奥运会奖牌榜排名只与它前三届的奥运会奖牌榜排名有关，且相关系数的大小与两届奥运会之间相差的届数有关。相差的届数相同的奥运会之间的相关系数也大致相当。于是，取上表中相差 1 届、2 届和 3 届的奥运会间相关系数的平均值来作为前 3 届奥运会对于 31 届奥运会的相关系数。于是得到相关系数向量  $r = (0.698, 0.539, 0.440)$ 。

### 三、利用时间序列和灰色预测模型对奥运会奖牌数进行预测

客观世界中既有大量已知信息又有大量未知信息，相互交织。如果将已知信息称为白色信息，未知或非确知信息称为黑色信息，既含有已知信息又含有未知或非确知信息的系统称为灰色系统。这些系统仅具有以时间为序列的数据，因此在研究时可将时间序列转化为微分方程，建立抽象系统发展变化的动态模型。

如果一个时间序列 $x_t$ 的变化受到自身变化的影响，那么我们就可以用时间序列模型来描述它，如果此序列是平稳的，我们可以用ARMA(p,q)模型来描述：

$$\begin{cases} x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \\ \phi_p \neq 0, \theta_q \neq 0 \\ E(\varepsilon_t) = 0, Var(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E(x_s \varepsilon_t) = 0, \forall s < t \end{cases} \quad (1)$$

如果经单位根检验此时间序列不平稳，但经过差分后平稳，我们可以用ARIMA(p,n,q)描述。

下表是选定的 16 个国家历届奥运会的奖牌总数。

	美国	德国	法国	英国	匈牙利	澳大利亚	意大利	韩国
1	19	13.5	11	6.5	5	2		
2	50	8	85	36.5	5	6	4	
3	190	13		2	4			
4	47	12	19	130	9	4	4	
5	61	25	14	41	8	6	6	
7	95		43	43		3	23	
8	99		31	34	9	6	16	
9	56	31	21	20	9	4	19	
10	93	21	19	16	15	5	36	
11	56	80	19	14	16		22	
14	84		29	21	27	13	29	
15	76		18	11	36	11	21	
16	74	26	14	24	26	35	25	
17	71	42	5	20	21	22	32	
18	90	50	15	18	22	18	27	
19	107		15	13	32	17	16	
20	94		13	18	35	17	18	
21	94		9	13	22	9	13	6
22			14	21	32	9	15	
23	156		28	37	0	24	32	19
24	94		16	24	22	14	14	30
25	108	82	29	20	30	27	18	39
26	91	65	37	15	21	41	35	27
27	97	56	38	28	17	52	34	28
28	103	48	33	30	17	49	32	30
29	110	41	40	47	8	46	27	31
30	104	44	34	65	17	35	28	28
31	121	42	42	67	15	29	28	21
	中国	俄罗斯	加拿大	瑞典	芬兰	荷兰	日本	巴西

3			5					
4			13	25	5			
5			8	58	26			
7			8	62	34	11	2	3
8			4	29	37	10	1	
9			15	25	25	17	5	
10			15	23	23	7	18	1
11			9	20	19	17	18	2
14			3	44	20	16		1
15			3	34	20	5	9	2
16			6	19	15	15	19	1
17			1	6	5	3	18	3
18			4	8	5	10	26	1
19			5	4	4	7	25	3
20			5	16	8	5	29	2
21			10	5	16	5	25	2
22				12	8	3		4
23	32		44	19	12	13	32	8
24	28		10	11	4	9	14	6
25	54		18	12	5	15	22	3
26	50	63	22	8	4	19	14	15
27	59	88	14	12	4	25	18	12
28	63	92	12	7	2	22	37	10
29	90	73	18	5	3	16	26	12
30	87	82	18	7	3	20	38	17
31	70	56	22	11	1	19	41	17

表3-1: 16个国家1-31届奥运会奖牌总数

根据文献[8]的研究,金牌数的东道主效应为11.31%,整体实力的东道主效应为11.71%,所以我们根据历届奥运会的东道主情况,将表一中的数据进行处理,消除东道主效应,东道主效应,是指体育比赛中主办地队伍会因为东道主的各方面优势取得出乎平常的好成绩的一种现象。此效应体现的甚为明显的是当今的两大比赛:奥运会赛场和足球世界杯比赛,历届比赛表明,东道主效应让主队获利甚多。所得到的数据仍然在表一中,用红色字体表示。下面以美国历届奥运会成绩为例,建立时间序列预测模型:

首先检验时间序列的平稳性,利用SAS统计软件,得到的检验结果为:(SAS程序见附录三)

Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F Zero
Mean	0	-2.3752	0.2809	-0.94	0.2986		
	1	-0.5135	0.5566	-0.32	0.5609		
Single Mean	0	-27.9829	<.0001	-5.74	0.0002	16.70	0.0010
	1	-43.6213	<.0001	-5.07	0.0004	13.12	0.0010
Trend	0	-31.5200	0.0001	-6.33	0.0002	20.14	0.0010

1      -67.9621   <.0001      -6.03    0.0003    18.33    0.0010

表 3-2: 单位根检验

时间序列模型是描述时间序列统一性的一种常用方法, 在预测未来事件时, 通过该事件的历史数据揭示事假未来的发展规律, 从而对该事件的未来做出预测。时间序列模型强调时间因素在预测中的主体作用, 暂不考虑外界因素的影响。所以在模型的建立中我们假设奥运会奖牌榜只与历届奥运会成绩相关, 暂不考虑其他因素的影响。由以上检验可知, 序列非平稳, 但消除线性趋势项后序列平稳。利用SAS建立的美国队奥运奖牌的ARIMA(p, n, q)模型为:

Minimum Information Criterion							
Lags	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5	MA 6
AR 0	5.065745	5.068902	5.074657	.	-34.995	.	.
AR 1	5.151372	5.131413	5.127452	.	-34.995	.	.
AR 2	4.00934	3.827098	-31.0876	.	-34.995	.	.
AR 3	-31.8525	-29.2298	-35.8523	.	-34.995	.	.
AR 4	.	.	.	.	-34.995	.	.
AR 5	-35.1634	.	-37.796	.	.	.	.
AR6	.	.	.	.	.	.	.

Error series model: AR(12)  
Minimum Table Value: BIC(5, 3) = -37.796

表3-3: ARIMA(p, n, q)模型确定

由上表, Minimum Table Value: BIC(5, 3) = -37.796, 选取BIC大。可知, 原始数据一阶差分后可建立ARMA(5, 3)模型, 即

$$x_t = 1 + 1.61912x_{t-1} + 2.13609x_{t-2} + 2.01795x_{t-3} + 1.0923x_{t-4} + 0.50015x_{t-5} + 0.93383\varepsilon_{t-1} + 0.98672\varepsilon_{t-2} + 0.6989\varepsilon_{t-3}$$

美国队下一届奥运会奖牌数的预测值及95%的置信区间如下:

Forecasts for variable x				
Obs	Forecast	Std Error	95% Confidence Limits	
28	114.3228	43.8989	80.9744	155.6106

表3-4: 美国队下一届奥运会奖牌数的预测

表四给出了美国队下一届奥运会奖牌数的预测值为 114.3228, 近似为 114 块, 95%的预测区间为 (80.9744, 155.6106)。

利用相同的方法可以得到其它 15 个国家下一届奥运会的时间序列预测方程、预测值。

法国:  $x_t = 1 + 0.47762x_{t-1} + 0.27179x_{t-2} + 0.16336x_{t-3} - 0.10356x_{t-4} - 0.23679x_{t-5}$

英国:

$$x_t = 1 + 0.70398x_{t-1} + 0.14323x_{t-2} + 0.06457x_{t-3} + 0.24604x_{t-4} + 0.31203x_{t-5} + 0.18806x_{t-6}$$

匈牙利:

$$x_t = 1 - 0.58446x_{t-1} + 0.33758x_{t-2} + 0.2274x_{t-3} + 0.10934x_{t-4} - 0.28107x_{t-5} - 1.33823\varepsilon_{t-1} + 0.91907\varepsilon_{t-2}$$

澳大利亚:

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。

如要下载或阅读全文，请访问：

<https://d.book118.com/727025042005006060>