

Logarithm Model Prediction for 2008 Beijing Olympic Games

Jiefu Wang, Yong Jiang

Department of Mathematics, Nanjing University of Science and Technology Nanjing, 210094, P. R. China

(Received August 10 2007, accepted November 2 2007)

Abstract. Considering the result of a sport event regarding years as a time serial, it can be described by a model of time serial including certain model and uncertain model. In this test we put forward logarithm model to predict sports results. According to the method provided in this text we carried on "prediction" of man's track events of 2004 Olympic Games and the "prediction" accuracy came to above 97%(average 99.1647%).At last, we built up logarithm model and carry on the error test. The result shows that logarithm model prediction can be believable and can be use to predict sports result.

Keywords: Olympics data, data pretreatment, logarithm model, gray model

1. Preface

We have used gray model to predict the Olympic Games results (reference [4]).At that time we discovered that prediction result isn't very ideal. So, in this test, we carry on a prediction model which is an improvement of gray model.

First, consider of the athletics facilities perfect day by day, people's body gradually raises and the continuously development of athletics science and technology, the Olympic Games result is toward to more excellent; Second, because of the physiology limit of human body, the Olympic Games result will tend to steady gradually. So we carry on logarithm model to predict the Olympic Games result.

2. The data pretreatment

The Olympic Games result's data has different data format and units. Sometimes the data may be very small, so the error may become bigger. Therefore, before carry on calculation, all the data about time may be converted into second as unit.

3. Model design and parameter compute

According to the analyses above, we adopt the model

$$y = a \ln x + b$$

Where x is the sequence of the years in which the Olympic Games hold. y is athletic result sequence. a, b are parameters.

Underneath we will solve a and b with the method of least squares.

Suppose we use a line polynomial

$$p(x) = a \ln x + b \quad (1)$$

to fit the data (x_i, y_i) . where x_i, y_i are the elements of the year sequences and the sports result sequence. At the point x_i the error is

$$R_i = p(x_i) - y_i$$

The method of least squares' thought is that find out a, b that make the error square's sum is least. Because

$$\Phi = \sum_{i=1}^n R_i^2 = \sum_{i=1}^n [p(x_i) - y_i]^2 = \sum_{i=1}^n (a \ln x_i + b - y_i)^2$$

So we make

$$\begin{cases} \frac{\partial \Phi}{\partial a} = 0 \\ \frac{\partial \Phi}{\partial b} = 0 \end{cases}$$

And solve this equation can get the value of a 、 b . So we get polynomial (1).

Disperse polynomial (1) we can got the logarithm model

$$p(x_i) = a \ln x_i + b, i = 1, 2, \dots, n \quad (2)$$

4. Prediction and error test

From formula (2) we can know:

When $i \leq n$, from the error between model value and concrete value, we can estimate the model.

When $i > n$, we can use it to predict the sports value.

Considering the precision, we give out the following estimate method (reference [3]).

Suppose e_i is the absolute error, namely

$$e_i = \frac{\left| \overset{\rightarrow}{p(x_i)} - y_i \right|}{y_i}, i = 1, 2, \dots, n$$

Also suppose e is the average value of e_i , Y is the average value of y_i , then $C = \frac{S_1}{S_0} = \frac{\frac{1}{n} \sum_{i=1}^n (e_i - e)^2}{\frac{1}{n} \sum_{i=1}^n (y_i - Y)^2}$

$$P = P\{|e_i - e| < 0.6745S_0\}$$

So we can carry on the accuracy grade of the prediction model by the table 1.

Table 1 accuracy grade form

Precision scale	P	C
Good(first scale)	>0.95	<0.35
Preferably(second scale)	>0.80	<0.50
Qualification(third scale)	>0.70	<0.65
Can(fourth scale)	>0.7(or ≤ 0.70)	≥ 0.65 (or <0.65)
Failed	≤ 0.70	≥ 0.65

We use original data of Olympic Games before 2000 to “predict” the result of 2004 Olympic Games by logarithm model (table 2). And compared with other kinds of prediction methods (table 3). The result express that the result we got is the best from various aspects.

Table 2 the 2004 Olympic Games "prediction" and the error

Project(men's)	Actual result	Predict result	error
100 m	9.85	9.9373	0.8864%
200 m	19.79	19.78	0.0503%
400 m	44	43.671	0.7474%
800 m	104.45	103.58	0.8317%
1500 m	214.18	215.51	0.6227%
10000 m	1625.1	1633.5	0.5180%
110 m hurdle	12.91	12.98	0.5411%
400m hurdle	47.63	47.101	1.1099%
3000m handicap	485.81	485.81	0.0002%
long jump	8.59	8.6357	0.5315%
high jump	2.36	2.3885	1.2095%
hop step and jump	17.79	17.888	0.5526%
Shot	21.16	21.85	3.2590%
discus	69.89	69.096	1.1361%
hammer	82.91	83.438	0.6368%
4×100 relay race	38.07	37.581	1.2844%
4×400 relay race	175.91	176.41	0.2823%

5. 2008 Beijing Olympic Games prediction

At last, according to the logarithm model this text established, we predicted the result of Beijing Olympic Games (table four). Here, we just consider the data before 2004.If in consideration of more other information, its predict accuracy will be higher.

6. Conclusion

This text only used the logarithm model in sports result prediction. Actually, the logarithm model can apply to predict other problems, such as the budget index predict, student's height predict and the residents live level predict etc.

Table 3 comparison of various predict methods

		This text	Linear regression	Accumulate Linear regression	Square regression	Cube regression	Gray model
$\sum e_i^2$		0.00203	0.004652	0.005278	0.00551	0.0133	0.003078
$\frac{1}{17} \sum e_i $		0.84%	1.25%	1.38%	1.23%	1.97%	0.98%
Error higher than this text's number and percent			12/70.59%	11/64.71%	10/58.8%	10/58.8%	10/58.82%
The max error		3.26%	4.18%	4.19%	5.00%	7.13%	3.71%
Error higher than	2%	5.88%	17.65%	23.53%	17.65%	41.18%	11.76%
	1%	29.41%	47.06%	64.71%	41.18%	58.82%	47.06%
	0.50%	82.35%	64.71%	76.47%	64.71%	70.59%	64.71%

7. Reference

- [1] Yong Jiang etc., the gray model method of sport result predict. *the athletics statistics*, 1987.12
- [2] Yong Jiang etc., the sports result predict by regression method , the second public lecturer academic athletics science reads out thesis, 1987.12
- [3] Xuemeng Wang etc, Gray system analysis and practical computational procedure, Huazhong university of science&technology publishing company,2001.9
- [4] Y. Jiang, J. L. Li, J. F. Wang, F.Y. Hua. Gray Models for the Prediction of Results in Track and Field Events. *International Journal of Sports Science and Engineering*.2007, 1(3)

Table 4 The prediction result of 2008 Beijing Olympic Games (i=11)

Sports Project(men's)	$y = a \ln x + b$		Precision prediction index		Precision scale	2008 prediction
	a	b	p	c		
100 m	-0.08783	10.13	0.90909	0.0031024	second	9.9115(s)
200 m	-0.19908	20.259	1	0.00072134	first	19.765(s)
400 m	-0.4918	44.919	1	9.41E-05	first	43.697(s)
800 m	-0.5743	105.14	1	1.78E-05	first	103.71(s)
1500 m	-1.1215	217.93	1	7.40E-06	first	215.14(s)
10000 m	-41.708	1731.8	1	6.55E-08	first	1628.1(s)
110 m hurdle	-0.27849	13.633	1	0.0003275	first	12.941(s)
400m hurdle	-0.78256	49.088	1	4.87E-05	first	47.143(s)
3000m handicap	-13.507	518.2	1	9.79E-07	first	484.64(s)
long jump	0.13721	8.2971	0.90909	0.0051079	second	8.6381(m)
high jump	0.087737	2.1722	0.36364	0.0081	forth	2.3902(m)
hop step and jump	0.41538	16.872	1	0.00041235	first	17.904(m)
Shot	0.54276	20.405	1	0.00054319	first	21.753(m)
discus	3.2028	61.581	1	2.32E-05	first	69.54(m)
hammer	5.5326	70.062	1	1.17E-05	first	83.809(s)
4×100 relay race	-0.4912	38.861	1	6.10E-05	first	37.64(s)
4×400 relay race	-1.7066	180.4	1	9.33E-06	first	176.15(s)