Available online www.jocpr.com

Journal of Chemical and Pharmaceutical Research, 2014, 6(4):1113-1119

Research Article

ISSN : 0975-7384
CODEN(USA) : JCPRC5

The Chinese top 110m hurdle athletes performance regression research based on MLR and GRA

Du Jingtao* and Wang Lei

Shijiazhuang Posts and Telecommunications Technical College, Shijiazhuang, China

---

ABSTRACT

Since Liu Xiang won the 28th Olympic Men's 110m hurdles champion, Chinese athletes' hurdle technique research has become a hot topic. In this paper, we focus on Chinese top 110m hurdle athletes' special features, by using multiple linear regression to analyze the correlation between Chinese top 110m hurdle athletes' performance and 4 factors (the number of 98% speed striding hurdles, the whole 10 hurdles step soaring time, the whole 110m hurdles' three pace running time, the number of striding hurdles before the Max speed turn down). Further on, we use grey relational analysis to find out which factor is dominant in the performance.

Key words: Multiple Regression analysis, Grey Relational Analysis, 110m hurdler, technical feature

---

INTRODUCTION

Men's 110m hurdle is one of the track and field events which required higher technique and stronger physical quality. Since Liu Xiang won the 28th Olympic Men's 110m hurdle champion, Chinese athlete's hurdle technique has won more and more attention from all over the world, and the related research has become a hot topic [1].

In 2009, Zhang YaPing, Cheng Hui published "The analysis of technique of the world-class men 110m hurdle athlete" indicated that the performance was dominated by average three pace running time [2]. In 2011, Guangzhou Sport University's Deng Wan-Jin, Liu Yong-Dong published "The GRA correlation analysis of world first-class 110m hurdles athletes' ability indicator and performance" in the Journal of Hebei Institute of Physical Education, mainly used GRA model to analyze some indexes (like speed lost hurdles soaring, hurdler striding time, the ability of integrate running and striding, and so on), and found out that the ability of integrate running and striding was the most important[3]; In 2012, Hebei University's Tang Lei published "The analysis of our country first-class 110m hurdle athletes' features of hurdler striding time", mainly pointed out the lacks of the hurdle technique[4]. In 2013, College of Applied Science and Technology of Hainan University's Tang Hua published "The Technique Features of the World's Top 110m Hurdlers in Whole Courses" in Sichuan Sports Science, mainly discussed the technique features in the new period [5-6]. In 2014, Huanggang Normal University's Bing Zhang published "110m Hurdles phased performance significance research based on SPSS regression analysis and GRA model" mainly put attention on four stages of whole 110m hurdles [7-8]. And in the same year, Xiao Nianle, Deng yanfang published "The analysis of 110m hurdle competitive capacity" [9].

In this paper, Based on the former studies and the interview results from Chinese first-class coaches, like Sun Hai-Ping and so on, we chose four factors, X1: the number of 98% speed striding hurdles (hereinafter referred to as number of hurdles), X2: the whole 110m hurdles' hurdle step soaring time (hereinafter referred to as hurdle step soaring), X3: the whole 110m hurdles' three pace running time (hereinafter referred to as three pace running), and X4: the number of striding hurdles before the Max speed turn down (hereinafter referred to turning point of Max speed). We use multiple linear regression theory to build regression equation, analyze the significance of the regression equation and coefficients while using GRA theory to analyze the four factors' position on upgrade 110m hurdles performance. Finally, we hope our research is useful to find out the feature of Chinese top 110m hurdle athletes' technological structure, and provide a reference for scientific training.

Data measurement method

For measure the original data, we put a JVC-9800 digital video(the shooting level is 125cm, the frame rate is 50fps) on the middle of the 110m hurdle track to track the whole running process like figure 1 shows. And using EIMG64PN-I imager to calculate and record the observable variables.

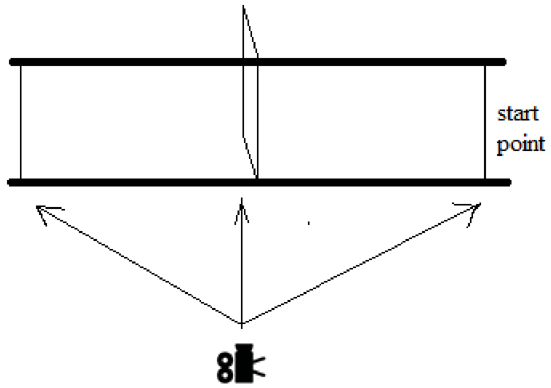

Figure 1: The DV shooting position map

Multiple linear regression analysis method

Multiple linear regression analysis is a common statistical method to analyze the linear relationship between one dependent variable and multiple independent variables. Though the observable variables, MLR can build regression equation, which could reveal the correlationship between dependent and independent variables. Now MLR has been widely applied to natural science, social sciences and economics.

Based on the MLR theory, we know there are independent variable Y and dependent variables X_1, X_2, X_3, X_4 meet:

$$Y = B_0 + B_1X_1 + \cdots + B_4X_4 + \varepsilon \tag{1}$$

Among them，Y is observable random variable，X_1 , X_2 , X_3 , X_4 are observable independent variables, b_0 is undetermined intercept, B_1 , B_2 , B_3 , B_4 are undetermined model parameters, ε is unobservable random error. Plug n groups independent sample date (y_i, x_{i1}... $x_{i4,}$ i=1, 2… n,) into formula (1), we get such formulas:

$$\begin{cases} Y_1 = b_0 + b_1x_{11} + \cdots + b_4x_{14} + e_1 ; \\ Y_2 = b_0 + b_1x_{21} + \cdots + b_4x_{24} + e_2 ; \\ \vdots \\ Y_i = b_0 + b_1x_{i1} + \cdots + b_4x_{i4} + e_i ; \end{cases}$$

i = 1, 2... n (2)

e_i are random variables and meet $Nor(0,\sigma^2)$, and we can use matrix to express formula (2) as :

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad X = \begin{pmatrix} 1, X_{11}, \cdots, X_{14} \\ 1, X_{21}, \cdots, X_{24} \\ \vdots \\ 1, X_{n1}, \cdots, X_{n4} \end{pmatrix} \quad B = \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \end{pmatrix} \quad e = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

Among them $\sum (e_i)^2$ = min, and according to least square method, finally, we get formula (3):

$$Y = XB \tag{3}$$

Using SPSS software to work out the undetermined parameters and testing
Step 1: According to the result, the multiple regression equation is:

$$y = 3.025 + 0.172x_1 + 1.993x_2 + 0.252x_3 + 0.061x_4 \quad (4)$$

Step 2: Test total regression equation's significance with F-test:
(1) Put forward null hypothesis, alternative hypothesis and significance level α :

$$H_0 : B_1 = B_2 = \cdots = B_k = 0 \quad H_1 : At\ least\ one\ B_k \neq 0 \qquad \alpha = 0.01$$

(2) Run ANOVA on the SPSS software and the result shows on Table 1, we can see the Sig. <0.001, and Sig. < α obviously. Therefore we refuse H_0, and the regression equation is significant.

Table 1: ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.354	4	.338	716.939	.000[a]
	Residual	.002	5	.000		
	Total	1.356	9			

a Predictors: (Constant), turning point of Max speed, number of hurdles, three pace running, hurdles step soaring
b Dependent Variable: total time

Step 3: Test all Regression coefficients
We use t-test to test all regression coefficients, and

(1) Put forward null hypothesis, alternative hypothesis and significance level α :

$$H_0 : B_k = 0 \quad H_1 : B_k \neq 0 \quad \alpha = 0.01$$

(2) we use SPSS to do all the mathematical calculation, and the result shows on Table 2:

Table 2: Regression Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3.025	0.727		4.162	0.009
	number of hurdles	0.172	0.023	0.377	7.543	0.001
	hurdles step soaring	1.993	0.130	1.212	15.355	0.000
	three pace running	0.252	0.030	0.340	8.498	0.000
	turning point of Max speed	0.061	0.019	0.195	3.313	0.021

a Dependent Variable: total time

According to Table (2) the four variables, X_1's(number of hurdles) t-statistics is 7.543, X_2's(hurdles step soaring) t-statistics is 15.355, X_3's(three pace running) t-statistics is 8.498, X_4's(turning point of Max speed) t-statistics is 3.313, and according to our significance level $\alpha = 0.01$, we can find out:

(1) Sig.1=0.001< α , it means the independent variable "number of hurdles" regression coefficient is significant;
(2) Sig.2<0.001< α , it means he independent variable "hurdles step soaring" regression coefficient is significant;
(3) Sig.3<0.001< α , it means he independent variable "three pace running" regression coefficient is significant;
(4) Sig.4=0.021> α , it means he independent variable "number of hurdles" regression coefficient is not significant;

According to the above conclusions, the independent variables X_1(number of hurdles), X_2(hurdles step soaring), X_3(three pace running) have significant influence to the dependent variable Y, while the X_4's effect on the 110m hurdles performance also can not be ignored, but in our case, we set a higher strict significance level, so we have to remove X_4 from regression equation.

Step 4: Rebuild the regression equation.
From the above, we remove X_4 from the regression equation, use SPSS software to do the multiple linear regression progress again, and finally, the multiple regression equation is:

$$y = 5.068 + 0.139x_1 + 1.678x_2 + 0.206x_3 \quad (5)$$

The regression equation F-test result is shown as Table (3) and the Sig. <0.001, obviously, Sig. < α .

Table 3: ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.348	3	0.449	357.603	0.000[a]
	Residual	0.008	6	0.001		
	Total	1.356	9			

a Predictors: (Constant), three pace running, number of hurdles, hurdles step soaring
b Dependent Variable: total time

The T-test of all regression coefficients result is shown as Table (4) and we can find out: all regression coefficients are significant.

Table 4: Regression Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5.068	0.628		8.074	0.000
	Number of hurdles	0.139	0.034	0.306	4.153	0.006
	Hurdles step soaring	1.678	0.144	1.020	11.628	0.000
	Three pace running	0.206	0.043	0.278	4.819	0.003

a Dependent Variable: total time

We can also use SPSS software to calculate the Durbin-Watson statistic. The Durbin-Watson statistic is often used as a tool to show whether all residuals are mutual independence or not. When the Durbin-Watson statistic = 2 or close to 2, means all residuals are mutual independence. In this paper, The Durbin-Watson statistic is 2.161, means all residuals are mutual independence.

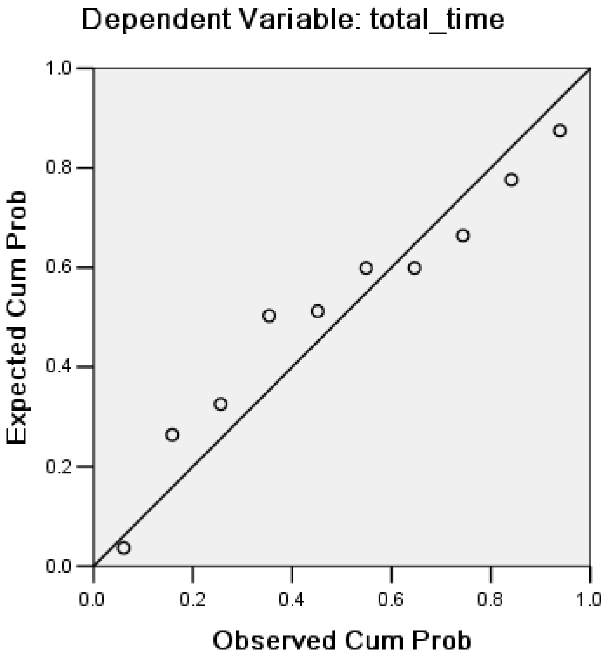

Figure 2: Normal P-P Plot of Regression Standardized Residual

From the Figure (1) we can find most of scatter points close to diagonal line, and it means the regression standardized residual is conforming to normal distribution.

By the conclusion of the multiple linear regression, we find out that the total time of 110m hurdles is correlated to X_1(number of hurdles), X_2(hurdles step soaring) and X_3(three pace running), X_4's(turning point of Max speed) is not significant like the others.

Chinese top 110m hurdles athletes' Grey relevancy analysis on 110m hurdles performance and four factors.
In 1979, Chinese professor Deng Ju-Long proposed Grey system theory, and in 1982, professor Deng published the first paper about Grey system theory "The Control Problems of Grey System" in System & Control Letter. Through several years' development, the theory has been widely applied in economics, military, sports, and other fields. Grey theory adaptation range is "small sample", "poor data information", "undefined" system, and it fits many aspects in sports research. In this paper, we regard 110m hurdles as a grey system project, using Grey relevancy analysis to

reveal which factor dominates 110m hurdles performance.

Step 1: Define reference series and contract series
In this case, we regard Y (total time of 110m hurdles) as reference series, and represented by X_0, regard X_1(number of hurdles), X_2(hurdles step soaring), X_3(three pace running), X4's (turning point of Max speed) as contract series. The detail is shown in Table (5).

Table 5: Original data series

NO.	X_0	X_1	X_2	X_3	X_4
1	12.9146	6.63	3.45	5.81	7
2	12.95	6.54	3.41	5.93	8
3	13.21	6.22	3.56	6.12	8
4	13.23	6.13	3.58	6.24	7
5	13.49	5.47	3.79	6.31	6
6	13.51	4.56	3.81	6.97	6
7	13.8	5.31	3.87	7.23	5
8	13.85	5.6	3.92	6.99	5
9	13.89	4.85	3.98	7.12	5
10	13.94	4.12	4.11	6.76	5

Step 2: Original data processing

Calculate each series' average values:

$X_0(t)$=13.47846 $X_1(t)$=5.543 $X_2(t)$=3.748 $X_3(t)$=6.548 $X_0(t)$=6.2

And then use average values to divide original data series, refer to Table (6):

Table 6: Processed Original data

X_0	X_1	X_2	X_3	X_4
0.958166	1.196103	0.920491	0.887294	1.129032
0.960792	1.179866	0.909819	0.90562	1.290323
0.980082	1.122136	0.94984	0.934637	1.290323
0.981566	1.105899	0.955176	0.952963	1.129032
1.000856	0.98683	1.011206	0.963653	0.967742
1.00234	0.822659	1.016542	1.064447	0.967742
1.023856	0.957965	1.032551	1.104154	0.806452
1.027565	1.010283	1.045891	1.067502	0.806452
1.030533	0.874977	1.0619	1.087355	0.806452
1.034243	0.74328	1.096585	1.032376	0.806452

Step 3: calculate grey correlation coefficient

From the grey correlation theory, we know the correlation coefficient formula is:

$$\xi_i(k) = \frac{\min_s \min_t |x_0(t) - x_s(t)| + \rho \max_s \max_t |x_0(t) - x_s(t)|}{|x_0(k) - x_i(k)| + \rho \max_s \max_t |x_0(t) - x_s(t)|} \qquad (6)$$

Among them, ρ is called resolution coefficient, and usually $\rho \in (0,1)$. ρ is used for waken the correlation coefficient distortion influence from the oversize $\max_s \max_t |x_0(t) - x_s(t)|$, so the value of ρ is depended on $\max_s \max_t |x_0(t) - x_s(t)|$ and assigned by the researcher. In common situation, ρ is often assigned closely to 0.5. In our case, ρ = 0.5.

Table (6) is the values of $|x_0(t) - x_s(t)|$:

Table 6: The absolute difference value

X_1	X_2	X_3	X_4
0.237937	0.037675	0.070872	0.170866
0.219074	0.050974	0.055172	0.329530
0.142054	0.030242	0.045446	0.310240
0.124333	0.026390	0.028603	0.147466
0.014026	0.010350	0.037203	0.033114
0.179681	0.014202	0.062107	0.034598
0.065891	0.008695	0.080298	0.217404
0.017282	0.018326	0.039936	0.221114
0.155556	0.031367	0.056822	0.224082
0.290963	0.062342	0.001866	0.227791

Form those values, we can find out:

$$\min_s \min_t |x_0(t) - x_s(t)| = 0.001866$$

$$\max_s \max_t |x_0(t) - x_s(t)| = 0.329530$$

$$\xi_i(k) = \frac{\min\limits_s \min\limits_t |x_0(t) - x_s(t)| + \rho \max\limits_s \max\limits_t |x_0(t) - x_s(t)|}{|x_0(k) - x_i(k)| + \rho \max\limits_s \max\limits_t |x_0(t) - x_s(t)|} = \frac{0.001866 + 0.5 \times 0.329530}{|x_0(k) - x_i(k)| + 0.5 \times 0.329530}$$

$$= \frac{0.166631}{x_{ij} + 0.164765} \tag{7}$$

Now we can calculate the $\xi_i(k)$, the result are as the Table 7 shows.

Table 7: Correlation coefficient table

X_1	X_2	X_3	X_4
0.413782	0.823113	0.707151	0.496470
0.434117	0.772374	0.757630	0.337108
0.543093	0.854486	0.792685	0.350798
0.576382	0.871706	0.861728	0.533678
0.931988	0.951553	0.825036	0.842084
0.483765	0.931070	0.734471	0.835817
0.722423	0.960632	0.679951	0.436014
0.915317	0.910101	0.814021	0.431822
0.520201	0.849588	0.751990	0.428526
0.365637	0.733711	0.999997	0.424477

Step 4: calculate correlation degree.

Correlation coefficient compares the correlation in all periods between contract series and reference series, so we should calculate correlation coefficient's average value as the quantities of the correlation between contract series and reference series. The correlation degree computational formula is:

$$\gamma_i = \frac{1}{10} \sum_{1}^{10} \xi_i(k) \tag{8}$$

By using formula (8), we can find out:

$$\gamma_1 = \frac{1}{10}(0.413782 + 0.434117 + 0.543093 + 0.576382 + 0.931988 + 0.483765 +$$
$$0.722423 + 0.915317 + 0.520201 + 0.365637) = 0.590670$$

$$\gamma_2 = \frac{1}{10}(0.823113 + 0.772374 + 0.854486 + 0.871706 + 0.951553 + 0.931070 +$$
$$0.960632 + 0.910101 + 0.849588 + 0.733711) = 0.865833$$

$$\gamma_3 = \frac{1}{10}(0.707151 + 0.757630 + 0.792685 + 0.861728 + 0.825036 + 0.734471 +$$
$$0.679951 + 0.814021 + 0.751990 + 0.99997) = 0.792466$$
$$\gamma_4 = \frac{1}{10}(0.496470 + 0.337108 + 0.350798 + 0.533678 + 0.842084 + 0.835817 +$$
$$0.436014 + 0.431822 + 0.428526 + 0.424477) = 0.511679$$

Step 5: Sort correlation degree

Obviously, $\gamma_2 > \gamma_3 > \gamma_1 > \gamma_4$, it means the first-class factors are hurdle step soaring and three pace running, among them, hurdle step soaring is primary; number of hurdles and turning point of Max speed are secondary.

CONCLUSION

Through the regression analysis process, the factor turning point of Max speed is not brought into the regression equation, it does not mean there is no correlation, but in this paper, we want to highline the Chinese top 110m hurdles athletes' features, so we set a very strict significant level. Finally, we find out: the Chinese top 110m hurdles athletes' secret of success is significantly correlated to number of hurdles, hurdle step soaring and three pace running.

The result of Grey correlation degree analysis shows that Chinese top 110m hurdles athletes' performance is dominated by the hurdle step soaring, and the others' correlation sequence is as follows: three pace running> number of hurdles> turning point of Max speed. And according to the correlation degree, we can easily divide those factors into two groups: the high correlative group and the low correlative group. In the former group we can find hurdle step soaring ($\gamma_2 = 0.865833$) and three pace running ($\gamma_3 = 0.792466$), which are mainly representative of technical aspects. While in the other group we can find number of hurdles ($\gamma_1 = 0.590670$) and turning point of Max speed ($\gamma_4 = 0.511679$), which are mainly representative of physical aspects. The difference of the correlation degree between two groups indicates that the technical advantage is the dominant factor that helps Chinese 110m hurdle athletes achieve excellent performance.

REFERENCES

[1] Cao QingJun; Cao SiYu, *Sports Culture Guide*, **2011**, (10), 33-36.
[2] Zhang Yaping; Chen Hui, *Journal of Gansu Lianhe University (Natural Sciences)*, **2009**, 23(5), 87-89.
[3] Deng Wanjin; Liu Yongdong, *Journal of Hebei Institute of Physical Education*, **2011**, 25(2), 62-66.
[4] Tang Lei, *Contemporary Sports Technology*, **2012**, 2(31), 20-21.
[5] Tang Hua; Deng Jiajia; Xiao Jianjun, *Sichuan Sports Science*, **2013**, (4), 63-66.
[6] Zhang B.; Zhang S.; Lu G. *Journal of Chemical and Pharmaceutical Research*, **2013**, 5(9), 256-262.
[7] Yujie Fan, *Journal of Chemical and Pharmaceutical Research*, **2014**, 6(2):261-268
[8] Bing Zhang. *Journal of Chemical and Pharmaceutical Research*, **2014**, 6(2), 649-659.
[9] Xiao Nianle; Deng Yanfang, *Sports World*, **2014**, (1), 82-83, 42.