Multi-factor Statistical Analysis of Influencing Factors for Elite Sports Development

Yong Jiang [+], Jianliang Li, Fengyou Hua and Jiefu Wang

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

Abstract. Sports developing depends on many relevant factors, such as local economy, education, population, geography, climate, GDP, culture, etc. In this paper, the achievement of nations in main games (such as Olympic Games) is used as dependent variable to indicate local sports developing level. 22 possible factors regarding sports developing level are considered as independent variable. By the main games data and linear regression method, we can find the relationship between the sport developing and the factors, directing the sport developing further.

Keywords: Statistics, Elite Sports, Economic Development.

1. Introduction

The development and prosperity of sport can promotes athletic sports prevalence and popularization. At the same time it's significant for school teaching and training. But how does athletics sports develop? Not only the selection, training, competition are important, but also economical development, science's education, population life quality and district environment characteristics also can't neglect. This paper carried on the statistical analysis with the mass data, theoretically discussing our country athletics sports development using the statistical result of various factors.

2. Basic thought and method

Based on the main games mass data and linear regression method, firstly, 22 models of one variable are created to express the relationship between sports developing level and each factor, respectively. Relevant coefficients and F-texts of models are very useful for understanding the effect of each factor as an independent variable. Secondly, all of 22 factors are classified to 6 groups, and 6 multivariable models are formed to describe the relationship between sports developing and each group. Thirdly, based on t-test, a stepwise statistical analysis is used to find prominent relative factors, and the final multi-element statistical model is obtained to open out the relationship between sports developing level and main factors. After finishing these work, we analysis the results and obtain concludes.

3. Indices selection

It's well-know that economy is the primary factor which affects athletic sports development. In additional, in our country rural area productivity level relatively is low so that the population quantity being engaged in sports is less, and stadium field facilities insufficiency. In view of these disadvantageous conditions, our country competitive sports give first place to the city. In a word, while we elect indices, we must consider national conditions synthetically.

We elect 22 indices, including two economic general situations indices, three sports general situations indices, five agricultural economic indices, four industrial economic indices, five education and population indices and three district characteristic indices. The indices' concrete significance is shown in Table 1.

[+] Corresponding author. *Tel.*: +86-25-84315878, *Fax.*: +86-25-84315878
  E-mail address: math@mail.njust.edu.cn.

Table 1 Significance of indices

Number	Groups	Tabs	Names	Unit
1	Economy	x_1	National income	Hundred million
	Economy	x_2	Per capita income	Dollar
2	Sports	x_3	Sports funds	Ten thousand
	Sports	x_3'	Per capita sports funds	Dollar
	Sports	x_4	Mass sports comprehensive value	Dollar
3	Agriculture	x_5	Agricultural society total output	Hundred million
	Agriculture	x_6	The per capita income of farmers	Dollar
	Agriculture	x_7	Agricultural total output	Hundred million
	Agriculture	x_8	Rural electricity	Hundred million kilowatt-hour
	Agriculture	x_9	grain total output	Ten thousand tons
4	Industry	x_{10}	Infrastructure investments	Hundred million
	Industry	x_{11}	Sanitation, sport and welfare investments	Hundred million
	Industry	x_{12}	Posts and telecommunications investment	Hundred million
	Industry	x_{13}	The total volume of retail sales	Hundred million
5	Education	x_{14}	Experts and engineers	Individual
	Population quality	x_{15}	University graduates	Individual
	Population movements	x_{16}	Percentage of immigrant population	%
	Population incomes	x_{17}	Per-capita wages of the person engaged in Sanitation, sport and welfare	Dollar
	Population quality	x_{18}	Percentage of illiteracy	%
6	District	x_{19}	Average temperature	Centigrade
	District	x_{20}	Average rainfall	Millimeter
	District	x_{21}	Average time of sunlight	Hours

4. Analysis and discussion

4.1. The relationship between sport developing and each factor, respectively.

Based on the simple linear regression model, we can obtain the relationship of sports developing and each factor. The results are shown in Table 2. By the simple regression analysis, we know the F-valve and

correlation coefficient R are the most important indices which weight the relativity between independent variable and dependent variable.

Using correlation coefficient R, in Table 2, we have obtained the interrelated data. If $R_{0.05} <| R |< R_{0.01}$, then x and y have prominent linear relationship, denote *, if $| R |< R_{0.01}$, then x and y have very prominent linear relationship, denote **, if $| R |< R_{0.05}$, then x and y have no prominent linear relationship. From Table 2, by the correlation coefficient, we obtain the result: x_1, x_2, x_3, x_5, x_6, x_7, x_8, x_{10}, x_{11}, x_{12}, $x_{13}, x_{14}, x_{15}, x_{16}, x_{18}, x_{20}$ and y have very prominent linear relationship respectively, x_4, x_{19} and y have prominent relationship respectively, but $x_3^{'}$, x_9, x_{17}, x_{21} and y have on prominent linear relationship respectively.

Table 2 The result of the simple regression analysis

Names of indices	Tabs	Regression equations	R	F-valve	relativity
National income	x_1	y=1.12809 x_1+38.12621	0.751	34.947	**
Per capita income	x_2	y=-0.212 x_2+159.148	0.509	9.435	**
Sports funds	x_3	y=0.141 x_3+(-98.874)	0.788	44.418	**
Per capita sports funds	$x_3^{'}$	y=4.412E-03 $x_3^{'}$+292.954	0.1438	0.571	
Mass sports comprehensive value	x_4	y=7.367E-02 x_4+67.805	0.428	6.05	*
Agricultural society total output	x_5		0.544	11.344	**
The per capita income of farmers	x_6	y=5.065E-03 x_6-149.8	0.654	20.265	**
Agricultural total output	x_7	y=1.133E-02 x_7+166.536	0.516	9.781	**
Rural electricity	x_8	y=0.733 x_8+182.741	0.536	10.8973	**
Grain total output	x_9	y=7.994E-03 x_9+237.638	0.320	3.065	
Infrastructure investments	x_{10}	y=8.92E-02 x_{10}-26.94	0.896	109.936	**
Sanitation, sport and welfare investments	x_{11}	y=1.444 x_{11}+206.9688	0.629	17.691	**
Posts and telecommunications investment	x_{12}	y=3.446 x_{12}+173.9446	0.775	40.525	**
The total volume of retail sales	x_{13}	y=0.178 x_{13}+7.873	0.802	48.827	**
Experts and engineers	x_{14}	y=8.053E-03 x_{14}+248.77	0.485	8.305	**
University graduates	x_{15}	y=1.667E-02 x_{15}+43.556	0.67	22.062	**
Percentage of immigrant population	x_{16}	y=0.631 x_{16}+131.7084	0.59	14.518	**
Average wages of the person engaged in Sanitation, sport and welfare	x_{17}	y=-4.853E-02 x_{17}+424.88	-5.3E-02	7.628E-02	
Percentage of illiteracy	x_{18}	y=-0.116 x_{18}+678.4	-0.537	10.922	*
Average temperature	x_{19}	y=2.411 x_{19}+28.966	0.439	6.443	*
Average rainfall	x_{20}	y=3.052E-02 x_{20}+85.988	0.554	11.93	*
Average time of sunlight	x_{21}	y=-1.595E-02 x_{21}+697.5	-0.291	2.49	

Remark: $F_{0.01}(1,27) = 7.68$ $F_{0.05}(1,27) = 4.21$ $R_{0.01}(27) = 0.47$ $R_{0.05}(27) = 0.367$

Similarly, by the F-test we also obtain the result: if $F_{0.05}(1,27) < F < F_{0.01}(1,27)$, then x and y have prominent linear relationship, denote *, if $F > F_{0.01}(1,27)$, then x and y have very prominent linear relationship, denote **, if $F < F_{0.05}(1,27)$, then x and y have no prominent linear relationship. By Table 2, we can get the same conclusions.

From Table 2, we can see: x_{10} (Infrastructure investments) has the closest linear relationship with spots developing, but it's only simple analysis, unilateral. And most factors of economy and industry have prominent linear relationships with sports developing. It's obvious that each factor of economy is important to sports developing, and simultaneity the sports funds, grain total output, sanitation, sport and welfare investments, average wages of the person engaged in Sanitation, sport and welfare and average time of sunlight of per year have no prominent linear relationships with sports developing. Thus, just considering the effect of individual factor, neither sports funds nor sanitation, sports and welfare investment is the leading factor. Sport developing is decided by the economic strength and industrial and agricultural development level of one nation. Accordingly, we must try our best to develop economy and enhance economic strength, further promoting sports developing. Although there are exceptions, the fact is only when the economy is strong can sport develops further.

In addition, we think per-capita sports funds is more power than the total sports funds, but fact is opposite. It tells us that when we do decision-making, we must depend on science, avoiding blindness.

4.2. The relationship between sport developing and each group, respectively.

Above we consider the factor respectively, but in practice all the factors interact. So all of 22 factors are classified to 6 groups, and 6 multivariable models are formed to describe the relationship between sports developing and each group. By multiple linear regression analysis, we can obtain each group results, which are shown in Table 3.

Table 3 The result of multiple linear regression analysis for each group

Group	R	F-value	Regression equation	t-test value
Economic integration	0.689	28.82	y=-64.4461+1.0093 x_1 +0.151 x_2	$t_3 = 5.626, t_4 = 2.156$
Sport	0.619	21.12		$t_3 = 5.626, t_4 = 2.156$
Agriculture	0.828	22.917	y=349.55-0.015 x_5 +0.0095 x_6 +0.0534 x_7 +0.99 x_8 -0.0268 x_9	$t_5 = -18.73, t_6 = 7.2387$ $t_7 = 28.147, t_8 = 8.3916$ $t_9 = -12.42$
Industry	0.922	71.18	y=-24.77+0.00693 x_{10} -0.6985 x_{11} +3.467 x_{12} +0.1234 x_{13}	$t_{10} = 1.223, t_{11} = -5.348$ $t_{12} = 13.696, t_{13} = 9.798$
Education and population	0.64	8.18	y=-230.311+0.00085 x_{14} +0.0083 x_{15} +0.367 x_{16} +0.3356 x_{17} -0.082 x_{18}	$t_{14} = 0.4069$ $t_{15} = 2.6742$ $t_{16} = 2.752$ $t_{17} = 2.93$ $t_{18} = -3.04$
District	0.331	4.12	Y= - 340.75 + 0.785x_{19} - 0.0338x_{20} + 0.0134 x_{21}	$t_{19} = 0.873$ $t_{20} = 3.75$ $t_{21} = 1.496$

$F_{0.01}(2,26) = 5.53$ $F_{0.01}(3,25) = 4.68$ $F_{0.01}(4,24) = 4.22$ $F_{0.01}(5,23) = 3.94$

$t_{0.01}(26) = 2.779$ $t_{0.01}(25) = 2.787$ $t_{0.01}(24) = 2.797$ $t_{0.01}(23) = 2.807$

Via the multiple linear regression analysis, we know that the relativity of regression equation is reflected

by correlation coefficient R and the value of F-test. Obviously, the sequence of influence is: Industrial economy (R=0.92)>Agricultural economy (R=0.828)>Economic integration (R=0.689)>Education and population (R=0.64)> Sport(R=0.619)>District(R=0.331), and test of the regression equation of District is failing, namely $F < F_{0.01}(3,25)$.

The result above also indicates that economy is the leading factor which influences sports developing. And all the indices about industry, agriculture and economy have prominent linear relationship. But we also find that some indices have prominent linear relationship in the simple linear regression and not here; even have not prominent effect, like Infrastructure investments. In the simple linear regression, the Infrastructure investment is the most correlative with sport developing, but is not prominent here. This proves the previous judgment further, namely we can not only consider factor respectively, must research them synthetically.

The following work is to consider all the factors together. We also use the multiple linear regression analysis, obtaining the regression equation and test result.

4.3. The relationship between sport developing and all the factors

Considering 22 possible factors regarding sports developing level together, we do multiple linear regression analysis, and the result is shown in Table 4.

Table 4 The result of multiple linear regression analysis for all factors

Regression equation	R	F-value	t-test value
$y=-704.49+1.054\,x_1$ $-0.086\,x_2 +0.0212\,x_3$ $-0.0075\,x_3' -0.000408\,x_4$ $-0.003698\,x_5 +0.00348\,x_6$ $+0.00866\,x_7 -0.6348\,x_8$ $-0.0288x_9 -0.00374\,x_{10}$ $-1.069\,x_{11} -0.3697\,x_{12}$ $+0.289\,x_{13} +0.01349\,x_{14}$ $-0.0065\,x_{15} +0.1712\,x_{16}$ $+0.0495\,x_{17} +0.00829x_{18}$ $+1.05947\,x_{19} +0.0081\,x_{20}$ $+0.018\,x_{21}$	R=0.992	F=34.56	$t_1 = 19.426$ $t_2 = -5.733$ $t_3 = 3.405$ $t_3' = -6.809$ $t_4 = 0.0655$ $t_5 = -10.9232$ $t_6 = 6.3298$ $t_7 = 10.9232$ $t_8 = -19.351$ $t_9 = -31.949$ $t_{10} = -1.596$ $t_{11} = -12.861$ $t_{12} = -2.30109$ $t_{13} = 36.165$ $t_{14} = 22.494$ $t_{15} = -7.247502$ $t_{16} = 4.443$ $t_{17} = 1.4979$ $t_{18} = 1.06268$ $t_{19} = 5.338$ $t_{20} = 4.0702$ $t_{21} = 9.111$

$t_{0.01}(6) = 3.707$ ($x_3, x_4, x_{10}, x_{12}, x_{17}, x_{18}$, can be deleted.) $F_{0.01}(22,6) = 7.35$

Since the number of data is the same as the number of independent variable, by the regression equation,

we obtain F-value and R being large, and R approximating 1. This indicates that 22 factors act on sports developing together, and effect is very obvious. Of course, we must find out main factors further. Using the correlation coefficient R, we can find the t-value of each coefficient in Table 4. There are six indices whose t-value is less than critical value. They are $x_3, x_4, x_{10}, x_{12}, x_{17}$ and x_{18}. These factors can be deleted. In order to find main factors, we make multiple linear regression analysis again with the residual factors, and the result is shown in Table 5. If testing the R and F-value, we find $x_3^{'}$ not satisfies conditions. Hence we delete $x_3^{'}$ and there are fifteen factors remained.

Table 5

Regression coefficient	R	F-value	t-test value
$\beta_0 = -578.88$			$t_0 = -77.035$
$\beta_1 = 1.389$			$t_1 = 32.98$
$\beta_2 = -0.161$			$t_2 = -13.757$
$\beta_3^{'} = -0.01038$			$t_3^{'} = -1.207$
$\beta_5 = -0.00505$			$t_5 = -19.351$
$\beta_6 = 0.00328$			$t_6 = 7.6868$
$\beta_7 = 0.0104$			$t_7 = 16.875$
$\beta_8 = -0.497$			$t_8 = -12.971$
$\beta_9 = -0.0289$	R=0.9906	F=78.660	$t_9 = -41.220$
$\beta_{11} = -1.171$			$t_{11} = -18.194$
$\beta_{13} = 0.245$			$t_{13} = 39.519$
$\beta_{14} = 0.0122$			$t_{14} = 26.233$
$\beta_{15} = -0.00564$			$t_{15} = -8.05$
$\beta_{16} = 0.1911$			$t_{16} = 6.3887$
$\beta_{19} = 0.968$			$t_{19} = 6.279$
$\beta_{20} = 0.0103$			$t_{20} = 6.723$
$\beta_{21} = 0.0165$			$t_{21} = 10.725$

$$F(16,12) = 3.97 \quad t_{0.01}(12) = 2.681 \quad \left|t_3^{'}\right| = \left|-1.207\right| < t_{0.01}(12) \text{ (delete } x_3^{'}\text{)}$$

Whether or not these factors have tangible effect? For this reason, we make multiple linear regression analysis again. The result is shown in Table 6. We know that the linear relativity of regression equation is remarkable, and the testing of correlation coefficient all pass. This indicates that sport developing is mainly determined by the fifteen factors in our country. We will analyze those factors as following：

4.3.1 The importance of agriculture

Although the influence of industry is larger than agriculture in second step, all the factors about agriculture remain in the last analysis, not being deleted for testing, and the t-values of x_5, x_7, x_8, x_9 is larger comparatively. This indicates that agriculture is still a key link in recent stage. Because china has a vast territory and a large population, our nation is still a representative agricultural country. Therefore, in order to promote sports level further and keep sport stable, sustained and coordinated development, the agriculture developing must not be ignored.

Table 6

Regression coefficient	R	F-value	t-test value
$\beta_0 = -578.876$			$t_0 = -79.787$
$\beta_1 = 1.388$			$t_1 = 34.128$
$\beta_2 = -0.165$			$t_2 = -14.6369$
$\beta_5 = -0.0053$			$t_5 = -21.169$
$\beta_6 = 0.00309$			$t_6 = 7.509$
$\beta_7 = 0.0105$			$t_7 = 17.725$
$\beta_8 = -0.467$			$t_8 = -12.62$
$\beta_9 = -0.0285$	R=0.99046	F=89.996	$t_9 = -42.147$
$\beta_{11} = -1.199$			$t_{11} = -19.292$
$\beta_{13} = 0.251$			$t_{13} = 41.88$
$\beta_{14} = 0.0123$			$t_{14} = 27.328$
$\beta_{15} = -0.00585$			$t_{15} = -8.64615$
$\beta_{16} = 0.179$			$t_{16} = 6.196$
$\beta_{19} = 0.975$			$t_{19} = 6.553$
$\beta_{20} = 0.0105$			$t_{20} = 7.007$
$\beta_{21} = 0.0162$			$t_{21} = 10.88$

$F_{0.01}(15,13) = 3.80 \quad t_{0.01}(13) = 2.65$

4.3.2 The relationship between sport and industry

Among the four factors about industry, two factors are deleted, but the rest have close relationship with sport developing. The sanitation, sport and welfare investments and the total volume of retail sales embody the concern of one nation for sanitation, sport and welfare and living standards of people. The total volume of retail sales is a reflection of production, besides it's also reflection of material civilization to a great extent. Only when people have proper material civilization can they have conditions to develop spiritual civilization. And sport developing begins this period, continual development and improvement.

4.3.3 The relationship between sport and education and population

Science and technology are first part of the productive forces, scientific and technological progress promotes development of productivity. Therefore education and science will exert far-reaching influence for sport developing. There are many athletes who cannot heighten their results further because of low educational level. Today science and technology quickly developing, sports competition means scientific and technological competition. Therefore in order to improving sports level, we must try our best to develop science and technology, enhance national quality.

Now our country is still a developing country; living standards of people is still improved further. There is a long way to catch up with the advanced level. Most people in China live in rural areas, whose quality is improving with the economic development. We must encourage more people participating in sport and contributing their share.

4.3.4 The relationship between sport and general economy

We know that economic strength and economic level have the most important influence to sports developing. We also can obtain a fact from Table 6. Namely, national income and average income all influence sports developing; it indicates the sport developing based on powerful economic strength, or the development been limited again.

4.3.5 The relationship between sport and district

In the second step, the factors about district hardly have influence on sports developing, but in overall analysis all the factors remain. This indicates that the regionally factors are not neglected. It is generally

know that all human activities are restricted by nature. Average temperature, average rainfall and average time of sunlight influence agricultural production, and the agriculture influences sports developing directly. Therefore while making sports decisions, we must consider not only man-made factors but also natural factors.

5. Four suggestions for sports developing

Through above analysis, we can get conclusions as following:

5.1 Our country is still an agricultural country, and agricultural economy is large part in national economy, besides, commercial economy is developing. Therefore，we must stabilize the agriculture and accelerate commercial economy rapid development, lying solid foundations for sports developing.

5.2 Accompanied economy developing, promoting educational level and population quality.

5.3 Making best use of regionally conditions, developing appropriate sports activities.

5.4 Enhancing economy strength, promoting sports level, making our athletic sport powerful step by step.

6. References

[1]   Xianlong Liu. Statistical Analysis of Sport and Economy. *Sports Science*. 1991, **1**

[2]   *Statistical yearbook of China*, Chinese Statistical Publishing.