Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 11, No. 3, 2024

Study on Optimal Training Methods for Teen-age 400-Meter Sprint Projects by Data Science

Zhicheng Fan

Shenzhen Middle School, Shenzhen Guangdong, China

Abstract: The research based on data science can provide more effective and personalized methods for the training of teen-age sprint events. This article analyzed the correlation between the consequent improvement in scores with gender, training methods, intensity, and cycles by collecting the key data of 400-meter running events. A multiple linear regression model was established to predict the running performance of teen-age athletes under different dependent variables, in order to provide references for the optimization of a teen-age training plan.

Keywords: 400-meter Running; Data Science; Correlation Analysis; Multiple Linear Regression Analysis; Prediction Model.

1. Introduction

The 400-meter distance running is a sprint event focused on speed and based on speed endurance. It is a periodic exercise of extreme intensity and was one of the earliest official Olympic athletic events [1]. To cultivate the physical fitness and competitive level of teen-age athletes, experts and scholars at home and abroad have conducted numerous theoretical and practical studies from the perspectives of physiology and training science. However, traditional training methods often rely on experience and intuition, lacking of scientific basis. In terms of how to effectively improve the running scores of 400-meter for teen-age athletes, there are no systematic approaches that have been formed yet. With the development of scientific data, sports training has also begun to attempt to use big data platforms and analysis tools for training optimization. The study on the optimal training of 400-meter running event for teen-age athletes by data science methods has certain auxiliary significance for improving running performance, and also provides more scientific, personalized and effective methods for teen-age trainers.

SPSS, one of the world's recognized statistical analysis software packages, is integrated general software for computer processing and statistical analysis [2]. It can be used in statistical operations, data mining, predictive analysis, decision support tasks, and etc. SPSS is widely applied to statistical analysis, data mining, predictive analysis, and decision support tasks with the usual purpose of the study of overall quantitative characteristics [3]. In addition, SPSS has also been used in the fields of sports training. According to Chen Jinshan et.al, SPSS is one of the commonly used tools for military sports scientific research and training data processing [4]. In general, the approaches of SPSS to achieve statistical analysis are divided into two categories. One of them is numerical calculation, that is, calculating the values of commonly used basic statistics to reflect the statistical characteristics of these data. And the other is graph rendering which can visually display the distribution characteristics of data through drawing common basic statistical graphs. This study adopts the method of data science to understand the impact of different factors on the improvement of running performance, such as gender, training methods, intensity, and cycles by collecting the key data from 400-meter running events. Based on the results of data analysis, a multiple linear regression model was established to predict the 400-meter running performance of teenagers. The obtained predicted results were expected to be helpful in the adjustment and optimization of training plans, thus realizing scientific and hierarchical training.

2. Methods and Materials

2.1. Search Strategies

Literature search was conducted on electronic databases including China National Knowledge Infrastructure (CNKI), VIP, and web of science with the keywords of "400-meter distance running of teenagers", "plyometrics training", "enhanced exercise", "the lower limbs explosive force" in accordance with requirements. Nearly 200 articles were carefully reviewed and selected.

Table 1. Mainstream training methods for teen-age athletes in 400-meter running [5-7]

Training plans	Training methods	Corresponding stages in 400-meter running
Strength training	Deep squat, half squat, power clean, snatch, barbell bench press, hurdle, stride jump, back step, split leg jump, split leg jump, mixed multi-hops, and etc.	Starting and sprint stage
Speed training	Listening to signals and running exercises, quick arm swing exercise, chopping of strides, trailer running, pimp running, and etc.	Full course stage
Speed endurance training	Variable speed running, interval running, repeat running, and etc.	Full course stage
Technical training	Single footed starting exercise, downwind running exercise, line collision exercise, and etc.	Intermediate stage
Corrective training	/	Full course stage

The theory and practice of competitive sports of 100-meter race indicated that physical fitness training, aimed at promoting coordination between tissue, organ and system function of teen-age organisms, is the foundation of technical training and improving sports performance. Therefore, modern scientific training must be comprehensive and efficient. The integrated system training methods for integrating physical training methods for 400-meter running were reviewed in Table 1. Results shows that the special training towards 400-meter running event contains strength training, speed training, speed endurance training, corrective training, and technical training, respectively corresponding to the starting, intermediate, sprint, and full course stages.

2.2. Questionnaire Investigation

The given questionnaire as shown in Table 2 was designed and formed trough several modifications combing domestic and foreign literature, with the above summary of training methods of 400-meter running. The specific situations of 400-meter running of teenagers were investigated from the coaches of sports institutions providing training for 400-meter running the by means of literature research and questionnaire. A total of 9 sports training institutions and 66 samples were counted for high school students aged 15-20 with over two years' experience in sprinting.

Table 2. Questionnaire for 400-meter training for teen-age athletes

For institutions: What are the training projects for 400-meter running in this institution? What are the corresponding training purposes, training skills, and training stages?							
Number	Training projects		Training purposes	Training skills	Corresponding stages		Cases requiring special clarification
1							
2							
For coaches: Please share with your students what training projects for 400-meter running and they have conducted and the results they have achieved?							
Gender	Age	Training projects	Average training days per week	Average training time per session	Scores before training	Scores after training	Cases requiring special clarification
1							
2							
3							
4							
5							

2.3. Mathematical Statistics

The resulting grades of teen-age athletes who had been trained were collected and the database was established statistically on the basis of the main training methods. Therewith, version 27.0 of SPSS software was applied in the correlation analysis for searching the correlations between the independent variable of improved running performance and the dependent variables of gender, training methods, training intensity, and training cycle. According to the analysis results, a coefficient closed to 1 suggests the strong positive correlation, and a coefficient closed to -1 inversely implies the strong negative correlation. If the value of coefficient is closed to 0, manifesting that there is no obvious correlation between the two. At the same time, the significance test is also indispensable to determine the statistical significance of the correlation coefficient. As a result, a p-value less than 0.05 is identified to be statistically significant, that is, the correlations between the improvement of training performance with gender, age, training method, training intensity, and training cycle are not accidental, but have a certain inevitability.

3. Results and Discussion

3.1. Analysis of the Effect on the Specialized Training

Table 3 shows the data results after organization from 66 samples. It turns out that the performance of teen-age athletes in the 400-meter running has been improved to varying degrees regardless of which training method has been suffered for them. Nevertheless, the same training program of different institutions have been carried out different training cycles in time and length. Due to the complex resources of samples, this data cannot intuitively reflect the correlation between performance improvement and various training methods.

3.2. Correlation Analysis of Training Performance Improvement

Correlation analysis usually refers to the utilization of SPSS software to measure the degree of correlation between two variable factors. In this study, correlation analysis was adopted to investigate the correlation between performance improvement and training projects (including strength training, speed endurance training, corrective training, speed training, and technical training), average weekly training time,

and gender. Pearson correlation coefficient method was employed to express the strength of the correlation. As shown in Table 4, the correlation value between performance improvement with speed endurance training is 0.256 while the significance level is 0.05, indicating a significant positive correlation between the two. On the contrary, the correlation value between performance improvement with average weekly training time is 0.254, and shows a significance of 0.05 level, suggesting that there is a significant positive correlation. In addition, the correlation values of 0.205, 0.147, and -0.056 together with p values of 0.098, 0.238, and 0.652 more than 0.05 hinted no correlation between performance improvement with speed training, corrective training, as well as gender.

Table 3. Comparison of 400-meter running performance of teen-age athletes before and after training

Number of institutions	Gender	Amounts of athletes	Training projects	Best performance before training	Best performance after training	Average improvement
1	Male	3	Strength training	52.57 s	50.06 s	4.73%
	Female	3		55.50 s	54.83 s	1.23%
2	Male	5	Technical training	49.02 s	48.16 s	1.76%
3	Female	2	Speed training	62.00 s	59.62 s	3.85%
4	Male	4	Speed training	56.50 s	53.71s	4.94%
5	Male	5	Strength training	56.86 s	52.28 s	8.04%
6	Male	10	Corrective training	51.90 s	51.82 s	0.15%
7	Male	26	Speed endurance training	53.18 s	51.64 s	2.88%
8	Male	1	Speed endurance training	54.00 s	49.86 s	7.66%
9	Male	7	Speed endurance training	48.91 s	48.57 s	0.70%

Table 4. Correlation analysis of performance improvement of teen-age athletes for 400-meter running

	Performance improvement	Speed endurance training	Corrective training	Speed training	Technical training	Average training time per week	gender	Strength training
Performance improvement	1							
Speed endurance training	0.256*	1						
Corrective training	0.205	-0.299*	1					
Speed training	0.147	-0.224	-0.134	1				
Technical training	0.122	-0.224	-0.134	-0.100	1			
Average training time per week	0.254*	0.284*	-0.041	-0.116	-0.024	1		
gender	-0.056	-0.202	0.039	0.109	-0.091	-0.163	1	
Strength training	0	0	0	0	0	0	0	1
* p<0.05** p<0.01								

3.3. Linear Regression Analysis of Training Performance Improvement

In addition to correlation analysis, SPSS can also be used for linear regression analysis. Linear regression analysis is a method of predicting the future value of a random variable based on the change of one or a group of independent variables.

Linear regression analysis is a method of predicting the future value of a random variable that is related to one or a group of independent variables based on their changes. Regression analysis requires the establishment of regression equations that describe the correlation between variables. The regression equation can be either univariate regression or multivariate regression depending on the number of independent variables. If the regression function is a linear function, the variables are said to be linearly correlated. The unary linear regression analysis consists of two variables, one of which is an independent variable, denoted by x. The other is that the dependent variable (the predictor) is represented by y.

It can be seen from the result of correlation analysis that the performance improvement is usually related to multiple variables. Consequently, the data characteristics in this study may be more suitable to be described by a linear regression model. Linear regression analysis was performed taking strength training as reference, Simultaneously that gender, training program, average weekly training time and best performance before training were consider to be independent variables. The independent variable was the improvement of training performance. The analyzed results were shown in Table 5. The R^2 value of 0.367 of the linear regression model means that gender, training methods, and average weekly training hours can explain 36.7% of the changes in performance improvement. The obtained F value of 5.704 and p value of 0.000 (< 0.05) passed the F-test, meaning that

gender, training methods, and average weekly training time would have an impact on the improvement of performance.

Table 5. Linear regression analysis of performance improvement of teen-age athletes for 400-meter running (n=66)

	Nonnormalized coefficient		Nonnormalized coefficient	t	p	Collinearity diagnostics	
	B	Standard error	Beta			VIF	Tolerance
Constants	-0.001	0.013	-	-0.079	0.937	-	-
Gender	0.005	0.010	0.050	0.468	0.642	1.080	0.926
Average training time per week	0.001	0.001	0.188	1.726	0.090	1.106	0.904
Corrective training	0.033	0.008	0.459	3.985	0.000**	1.239	0.807
Speed endurance training	0.028	0.007	0.511	4.054	0.000**	1.480	0.675
Speed training	0.033	0.010	0.374	3.349	0.001**	1.159	0.862
Technical training	0.030	0.010	0.344	3.055	0.003**	1.182	0.846
Strength training	0						
Adjusted R^2	0.367						
R^2	0.303						
F	F =5.704, p=0.000						
Value of D-W	1.041						
Dependent variable: Performance improvement							
* p<0.05** p<0.01							

Among them, the regression coefficients of gender and average weekly training time were 0.005 (t= 0.468, p=0.642 > 0.01) and 0.001(t=1.726, p=0.090>0.05), indicating that gender and weekly training time had no influence on performance improvement. In addition, the regression coefficients value of corrective training, speed endurance training, speed training, and technical training are 0.033 (t=3.985, p=0.000<0.01), 0.028 (t=4.054, p=0.000<0.01), 0.033 (t=3.349, p=0.001<0.01), and 0.030 (t=3.055, p=0.003<0.01), respectively. It suggests that all training methods consisting of corrective training, speed endurance training, speed training, and technical training showed significant positive impacts on the improvement of training performance.

In summary, corrective training, speed endurance training, speed training, technical training, and strength training will have the significant positive impacts on performance improvement of 400-meter running. Gender and average weekly training duration do not affect performance improvement. The final linear regression model formula is simulated as: Performance improvement=-0.001+0.005* gender+ 0.001* average weekly training time+0.033* corrective training+0.028*speed endurance training+ 0.033* speed training+0.030 * technical training.

3.4. Prediction of Training Performance Improvement

Table 6. Comparison between the actual and predicted running performance of 400-meter for teen-age athletes in a sports institution of Shenzhen

Number	Gender	Age	Training projects	Average training days per week	Average training time per session	Scores before training	Scores after training	Actual performance improvement	Predicted performance improvement	Standard error
1	Male	18	Speed endurance	3 days	2 hours	52.56 s	50.67 s	3.60%	3.8%	5.26%
2	Male	17	Speed endurance	3 days	2 hours	55.63 s	53.48 s	3.86%	3.8%	1.58%
3	Male	18	Speed endurance	3 days	2 hours	53.94 s	52.02 s	3.56%	3.8%	6.32%
4	Male	18	Speed endurance	3 days	2 hours	53.72 s	53.13 s	1.13%	1.0%	13.00%
5	Male	17	Speed endurance	3 days	2 hours	54.32 s	53.77	1.01%	1.0%	1.00%
6	Male	17	Speed endurance	3 days	2 hours	60.08 s	59.53 s	0.92%	1.0%	0.08%
7	Female	17	Speed endurance	3 days	2 hours	59.78 s	57.02	4.62%	4.3%	7.44%
8	Female	18	Speed endurance	3 days	2 hours	60.23 s	57.87	3.92%	4.3%	8.84%
9	Female	17	Speed endurance	3 days	2 hours	61.31 s	58.69	4.27%	4.3%	0.70%

The above linear regression model was used to predict and validate the performance improvement of 400-meter running by teen-age athletes. The training grades of 400-meter distance running from a sports training institution in Shenzhen before and after training are shown in Table 6. The calculated predicted values of performance improvement by substituting gender, age, training program, and average weekly training duration of teen-age athletes into the prediction model were compared with the actual values. The results shows that the average standard deviation is 4.91%, controlled within 5%. It can be considered that the prediction model is basically effective.

4. Conclusion

Big data is both a resource and a perfect research method. It is necessary to analyze massive amounts of data, extract them into knowledge, and then transform them into decisions and actions. It can provide a new path for optimizing sports training methods by fully utilizing data resources. It was proposed to conduct correlation analysis and linear regression models for analysis of the data of 400-meter running based on data science methods in this study. The main conclusions obtained are as follows:

(1) Any training method, including corrective training, speed endurance training, speed training, technical training, and strength training, has a significant improvement effect on the performance of 400-meter running for teenagers.

(2) Compared to correlation analysis, multiple linear regression analysis is more suitable for describing the data characteristics of 400-meter running performance improvement for teen-age athletes under different training methods.

(3) According to the results of multiple linear regression, corrective training, speed endurance training, speed training, technical training, and strength training would have a significant positive impact on performance improvement. However, gender and average weekly training duration do not have an impact on performance improvement. The linear regression model formula is simulated as: Performance improvement=-0.001+0.005*gender+0.001* average weekly training time+0.033*corrective training+0.028*speed endurance training+0.033*speed training+0.030 * technical training.

(4) The linear regression model was employed to predict the improvement of performance in 400-meter running for teenagers after training. The results showed that the predicted values were not significantly different from the actual measurements, and the error was controlled within 5%. This indicates that the model can be well applied to sports training for 400-meter running for teenagers, and can provide a certain theoretical basis for optimizing athlete training methods.

References

[1] Kai-Feng Z, Chong-De X. Correlative Analysis on 100-Meter Section Running and 400-Meter Racing Results [J].Journal of Beijing University of Physical Education, 2005.

[2] Heck R H, Thomas S L, Tabata L N .Multilevel and Longitudinal Modeling with IBM SPSS[J].Routledge, 2013.Nie N H ,HULL, Hadlai,et al. SPSS: Statistical package for the social sciences [J].acta chirurgica belgica, 1970.

[3] Jin-Shan C. Python-based big Data processing in military sports training [J]. Journal of military sports, 2020, 39(1): 4. (In Chinese)

[4] Yu-Ying Z. Further Research on Training of and Stamina in 400-Meter Running [J]. Journal of Hebei Polytechnic College, 2007.

[5] S. Rodríguez-Barbero, F. González-Mohíno, V. Rodrigo-Carranza, D. J. Santos-Garcia, D. Boullosa, J. M. González-Ravé. Acute effects of interval training on running kinematics in runners: A systematic review [J]. Gait & Posture, 2023, 19-26.

[6] D. Lum, F. Tan, J. Pang, T. M. Barbosa. Effects of intermittent sprint and plyometric training on endurance running performance [J]. Journal of Sport and Health Science, 2019, 471-477.

[7] Ying-Ying C. Effects of compound training of rapid expansion and contraction on explosive power of lower limbs of adolescent male sprinters [D]. Beijing Sport University, 2019. (In Chinese)