Study on the regression model of the relations between special quality and special results for high jump athletes

Yan Feng
Department of Physical Education, Zhengzhou University, Zhengzhou 450000, (CHINA)
E-mail: zb58@163.com

ABSTRACT

In order to study the relationship between the high jumpers’ special quality and its results, this article collects 14 world-class excellent jump athletes’ special quality and special performance on seven special items, as 100m run, standing triple jump, the run-up hand tough high, 4-6 strides approach high, after throwing the shot put, high grab the barbell and squat barbell. First of all, conduct correlation analysis of each special quality and its special performance. And the study found that there is certain correlation between the two. Secondly, conduct a principal component analysis of all special qualities and extract a principal component. Then build a regression model with the principal component data as independent variables and its special results as the dependent variables. This model reflects the relationship between the special qualities and special performances, and provides a reference for the daily training for coaches and athletes.

© 2013 Trade Science Inc. - INDIA

KEYWORDS

High jump;
Special quality;
Special performance;
Regression model.

INTRODUCTION

The high jump is a highly technical track and field event. With the development of the world track and field career, high jumpers have a very high level in their day-to-day training process and a growing number of athletes and coaches have started to pay attention to the training of special qualities. Therefore, in-depth study on the relationship of long jumpers’ various special qualities and corresponding performances possesses great realistic significance.

Relative experts and scholars from home and abroad have conducted a lot of research work on the relationship between special performance and special quality. For example, Laomin Li’s study shows that by strengthening high jump athlete’s training on special qualities as snatching, squatting, triple jumping, jumper’s relative special performance can be improved. Faping Chen, through statistical analysis of special achievements data and special qualities data and sequencing of long jumpers’ eight special qualities as: run between 30 meters, the 100 meters run and the standing triple jump, finds that factors as standing triple jump, run-up jump high and the run-up hand tough height have the greatest impact on athletes’ special performances. And previous studies are usually focused on the all various special qualities. Whereas, special qualities affecting high jump athletes’ special performances are of many kinds, and...
it is impossible to take into account all the special qualities by the High Jumpers in the day-to-day training. Therefore, identifying several major aspects affecting the high Jumpers’ special performances with an appropriate method, in order to further guide the training of coaches and athletes, is more realistic.

This study aims at conducting principal component analysis of various special qualities, achieving the drop-dimensional of special qualities. Then build a regression model with the main principal components and special achievements and analyze the importance degree of each principal component to special performance, in order to provide certain theoretical basis for high jumpers’ daily training.

**BRIEF INTRODUCTION OF THE METHOD**

**Correlation analysis**

The correlation analysis is a statistical analysis method that studies the correlation between the random variables, mainly by the correlation coefficient to determine the correlation. For two variables, \( x \) and \( y \), if the sample values are respectively \( x_i \) and \( y_i \), then the correlation coefficient of the two is:

\[
 r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \in [-1, 1]
\]

\[
 \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \quad \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i
\]

If the absolute value of \( r_{xy} \) is closer to 1, it demonstrates that the closer is the relationship between the two variables; and if the absolute value is closer to 0, it indicates that the relationship between the two variables is less closely.

**Principal component analysis**

The basic principle of the Principal Component Analysis:

Suppose that the number of indexes is \( p \), which are \( X_1, X_2, \ldots, X_p \). To find a set of independent composite indicators \( Z_1, Z_2, \ldots, Z_p \) that summarize the main information of the number \( p \) indexes, is to find a set of constants \( a_{11}, a_{12}, \ldots, a_{ip} \) (\( i=1,2,\ldots,p \)), that can realize the linear combination of the number \( p \) indexes, from a mathematical point of view:

\[
 Z_1 = a_{11} X_1 + a_{12} X_2 + \cdots + a_{1p} X_p
\]

\[
 Z_2 = a_{21} X_1 + a_{22} X_2 + \cdots + a_{2p} X_p
\]

\[
 \vdots
\]

\[
 Z_p = a_{p1} X_1 + a_{p2} X_2 + \cdots + a_{pp} X_p
\]

The linear combination can generalize the main information of the number \( p \) original indexes. For the convenience of description, introduce the following forms of matrix:

\[
 Z = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_p \end{bmatrix}, \quad A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{bmatrix}, \quad \Delta = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix},
\]

Then formula (1-1) can be expressed as \( Z = AX \) or

\[
 Z_1 = a_1' X \\
 Z_2 = a_2' X \\
 \vdots \\
 Z_p = a_p' X
\]

If formula \( Z_i = a_i' X \) satisfies \( a_i' = 1 \) and \( Var(Z_i) = \text{Max} \{ Var(a_i' X) \} \), then \( Z_1 \) is named as the first principal component of original indexes \( X_1, X_2, \ldots, X_p \). When \( Z_i \)“Zj, principal components \( Z_i \) and \( Z_j \) are unrelated, and \( Z_1 \) is the linear combination with the largest variance among all the linear combination of original indexes \( X_1, X_2, \ldots, X_p \); \( Z_2 \) is the linear combination with the largest variance among all the linear combination of original indexes \( X_1, X_2, \ldots, X_p \) except for \( Z_1 \). Sequence the indicators...
**Full Paper**

according to the variance size, and so on.

Basic steps of principal component analysis of the basic steps:

1) Standardization of the raw data. With application of Z-score method, the transformation formula is:

\[
Z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}
\]

In the formula, \(x_j = \frac{1}{n} \sum_{i=1}^{n} x_{ij}\), \(s_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2}\) \(i=1, 2, 3, \ldots, n\) \(j=1, 2, 3, \ldots, p\)

The mean of the transformed data is zero and the variance is 1.

The reason for original data standardization: in the principal component analysis, when determine the principal components by correlation coefficient matrix \(R\) (covariance matrix \(\Sigma\)), variables with larger variances \(\delta_j^2\) are often given priority to for consideration. In other words, this process is highly affected by the measurement scales of variables. Sometimes this will cause unreasonable results. In order to more objectively explain the connotation of the principal components, standardization of the original variable data must be conducted, avoiding the impact of the measurement units and magnitudes.

2) Determine the correlation matrix \(R\) of index data

The correlation coefficient matrix \(R\) of Variables (indicators) is the starting point of the principal component analysis. Measurement formula is:

\[
r_{ij} = \frac{1}{n-1} \sum_{i=1}^{n} \left( \frac{x_{ij} - \bar{x}_i}{s_i} \right) \left( \frac{x_{ik} - \bar{x}_k}{s_k} \right)
\]

\[
r_{ik} = \frac{1}{n-1} \sum_{i=1}^{n} Z_iZ_k
\]

And there is \(Rii=1\) and \(rik=rijk\).

3) Determine the Eigen value, Eigen vectors and the contribution rate of matrix \(R\)

The characteristic equation of matrix \(|\lambda I - R| = 0\), and \(\lambda_{g}(g=1, 2, \ldots, P)\) is the characteristic root of the characteristic equation. It is the variance of principal component \(Z\) and its value describes each principal component’s ability to summarize the original indexes’ information. And \(L\) represents a P-dimensional real vector. The vector \(Lg\) calculated from equations \([\lambda_i - \lambda]L = 0\) is the corresponding characteristic vector to characteristic root \(\lambda_i\). It is also the coefficient of each component on the new coordinates of standardized vector. \(Z = \left[ \begin{array}{c} z_{1i} \\ \vdots \\ z_{ni} \end{array} \right] \)

This formula shows each principal component explains the information of the original variables, that is variance contribution rate.

**Stepwise regression analysis**

Regression analysis is used to study the relationship between objective things. Built on the basis of a large number of observations and test of objective things, it is a statistical method to find those statistical regularity hidden in the outlook of uncertain phenomenon. In terms of theory and practice, multiple linear regression technique is a fairly classic quantitative analysis method.

The research objects of multiple linear regression model is affected by multiple factors: \(x_i, x_j, \ldots, x_n\). Assume that the relationship of various factors and \(y\) is linear and then the multiple linear regression models are:

\[
y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \epsilon_i
\]

Wherein, \(\beta_0, \beta_1, \ldots, \beta_p\) is the number unknown parameters, \(\beta_i, \ldots, \beta_p\) is called as the regression constant, \(\beta_0\) called as the regression coefficient, \(y\) is known as the dependent variable, \(x_i, x_j, \cdots, x_n\) is the general variables that can accurately measured and controlled, also called independent variables, \(\epsilon\) is Random error. There is the following hypothesis about the random error:

\[
D(\epsilon_i) = \sigma^2
\]

\[
E(\epsilon_i) = 0
\]

\[
\text{cov}(\epsilon_i, \epsilon_j) = \begin{cases} \sigma^2, & i = j \\ 0, & i, j = 1, 2, \ldots, n \\ \end{cases}
\]

To a practical problem, if the set number of observed data that can be obtained is \(n: (x_{ij}, x_{i2}, \ldots, x_{in}; y_i)\) and \(i = 1, 2, \ldots, n\). Then the linear regression model for-
The multiple linear regression model can be expressed as:

\[
\begin{align*}
    y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_p x_{1p} + \epsilon_1 \\
    y_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_p x_{2p} + \epsilon_2 \\
    &\vdots \\
    y_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_p x_{np} + \epsilon_n
\end{align*}
\]

In the form of matrix is \( y = XB + \epsilon \)

Wherein:

\[
\begin{bmatrix}
    y_1 \\
    y_2 \\
    \vdots \\
    y_n
\end{bmatrix} =
\begin{bmatrix}
    1 & x_{11} & x_{12} & \cdots & x_{1p} \\
    1 & x_{21} & x_{22} & \cdots & x_{2p} \\
    \vdots & \vdots & \vdots & \ddots & \vdots \\
    1 & x_{n1} & x_{n2} & \cdots & x_{np}
\end{bmatrix}
\begin{bmatrix}
    \beta_0 \\
    \beta_1 \\
    \vdots \\
    \beta_p
\end{bmatrix} +
\begin{bmatrix}
    \epsilon_1 \\
    \epsilon_2 \\
    \vdots \\
    \epsilon_n
\end{bmatrix}
\]

Estimate the regression parameters:

Estimate with the least squares method. Suppose the residual value between the true value and the estimated value of the model, then:

Wherein: \( \hat{y} = XB \cdot \hat{y} = XB \)

According to the requirements of the least squares method, there should be: \( E' E = (Y - \hat{Y})(Y - \hat{Y}) = \mathbf{M} \mathbf{n} \),

\[
E' E = (Y - XB)'(Y - XB) = \mathbf{M} \mathbf{n}
\]

Pole value principle: according to differential coefficient of matrix, determine the derivative of matrix \( B \), and suppose the value is 0, then:

\[
\frac{dE' E}{dB} = 2(Y - XB)'Y - 2(Y - XB)'X'B = 0
\]

The estimated value of the regression coefficient vector, \( B \) is: \( \hat{B} = (X'X)^{-1}X'Y \)

In accordance with the application conditions of the model:

1. The dependent and independent variables have a linear relationship;
2. The observations, \( Y_i (i = 1, 2, \cdots, n) \), are independent of each other;
3. The residual error \( \epsilon \) obeys the normal distribution with mean 0 and variance \( \sigma^2 \).

Select the training sample and test sample of the BP neural network modeling for modeling, and simulation samples are used to predict the multiple linear regression modeling. The predictive analysis sequence is modeling, inspection and prediction.

### STEPWISE REGRESSION MODEL OF THE RELATIONSHIP BETWEEN HIGH JUMPERS’ SPECIAL QUALITY AND SPECIAL PERFORMANCE

#### Research data

14 world elite male high jumpers’ data of specific performance and the special quality is collected, as shown in TABLE 1. Descriptive analysis of the 14 athletes’ special qualities and special performances was performed, and the results are shown in TABLE 2.

<table>
<thead>
<tr>
<th>Number</th>
<th>100m run /s</th>
<th>Standing triple jump /m</th>
<th>Run-up hand</th>
<th>4-6 steps run-up</th>
<th>After throwing shot /m</th>
<th>High grip barbell /m</th>
<th>Squat barbell /m</th>
<th>Specific Performance /m</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10.7</td>
<td>10.1</td>
<td>1.25</td>
<td>2.25</td>
<td>16</td>
<td>135</td>
<td>200</td>
<td>2.4</td>
</tr>
<tr>
<td>2</td>
<td>10.8</td>
<td>10.06</td>
<td>1.24</td>
<td>2.24</td>
<td>15.9</td>
<td>132.5</td>
<td>197.5</td>
<td>2.39</td>
</tr>
<tr>
<td>3</td>
<td>10.8</td>
<td>10.01</td>
<td>1.23</td>
<td>2.23</td>
<td>15.8</td>
<td>132.5</td>
<td>195</td>
<td>2.38</td>
</tr>
<tr>
<td>4</td>
<td>10.8</td>
<td>9.95</td>
<td>1.22</td>
<td>2.22</td>
<td>15.7</td>
<td>130</td>
<td>192.5</td>
<td>2.37</td>
</tr>
<tr>
<td>5</td>
<td>10.8</td>
<td>9.92</td>
<td>1.21</td>
<td>2.21</td>
<td>15.6</td>
<td>130</td>
<td>190</td>
<td>2.36</td>
</tr>
<tr>
<td>6</td>
<td>10.8</td>
<td>9.87</td>
<td>1.2</td>
<td>2.2</td>
<td>15.5</td>
<td>127.5</td>
<td>187.5</td>
<td>2.35</td>
</tr>
<tr>
<td>7</td>
<td>10.9</td>
<td>9.85</td>
<td>1.19</td>
<td>2.2</td>
<td>15.4</td>
<td>127.5</td>
<td>185</td>
<td>2.34</td>
</tr>
<tr>
<td>8</td>
<td>10.9</td>
<td>9.75</td>
<td>1.19</td>
<td>2.19</td>
<td>15.3</td>
<td>127.5</td>
<td>185</td>
<td>2.33</td>
</tr>
<tr>
<td>9</td>
<td>10.9</td>
<td>9.7</td>
<td>1.18</td>
<td>2.18</td>
<td>15.2</td>
<td>125</td>
<td>180</td>
<td>2.32</td>
</tr>
<tr>
<td>10</td>
<td>10.9</td>
<td>9.63</td>
<td>1.18</td>
<td>2.18</td>
<td>15</td>
<td>125</td>
<td>177.5</td>
<td>2.31</td>
</tr>
<tr>
<td>11</td>
<td>10.9</td>
<td>9.55</td>
<td>1.17</td>
<td>2.16</td>
<td>14.6</td>
<td>120</td>
<td>172.5</td>
<td>2.3</td>
</tr>
<tr>
<td>12</td>
<td>10.9</td>
<td>9.55</td>
<td>1.16</td>
<td>2.16</td>
<td>14.6</td>
<td>120</td>
<td>172.5</td>
<td>2.29</td>
</tr>
<tr>
<td>13</td>
<td>11</td>
<td>9.52</td>
<td>1.16</td>
<td>2.16</td>
<td>14.4</td>
<td>120</td>
<td>170</td>
<td>2.28</td>
</tr>
<tr>
<td>14</td>
<td>11</td>
<td>9.45</td>
<td>1.16</td>
<td>2.15</td>
<td>14.2</td>
<td>117.5</td>
<td>170</td>
<td>2.27</td>
</tr>
</tbody>
</table>
Table 2: The descriptive analysis result of the 14 athletes’ special qualities and special achievements

<table>
<thead>
<tr>
<th>Statistics</th>
<th>100m run /s</th>
<th>Standing triple jump /m</th>
<th>Run-up hand touch height /m</th>
<th>4-6 steps run-up height /m</th>
<th>After throwing shot /m</th>
<th>High grip barbell /m</th>
<th>Squat barbell /m</th>
<th>Specific Performance /m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>10.8643</td>
<td>9.7793</td>
<td>1.1957</td>
<td>2.195</td>
<td>15.2286</td>
<td>126.4286</td>
<td>183.9286</td>
<td>2.335</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.08419</td>
<td>0.21614</td>
<td>0.03031</td>
<td>0.03205</td>
<td>0.58366</td>
<td>5.43442</td>
<td>10.36425</td>
<td>0.04183</td>
</tr>
<tr>
<td>Maximum</td>
<td>10.7</td>
<td>9.45</td>
<td>1.16</td>
<td>2.15</td>
<td>14.2</td>
<td>117.5</td>
<td>170</td>
<td>2.27</td>
</tr>
<tr>
<td>Minimum</td>
<td>11</td>
<td>10.1</td>
<td>1.25</td>
<td>2.25</td>
<td>16</td>
<td>135</td>
<td>200</td>
<td>2.4</td>
</tr>
</tbody>
</table>

**Correlated analysis of special quality and specific performance**

With person correlation analysis, determine the correlation coefficient of each special quality and the specific performance, and the results is in Table 3.

As can be seen from Table 3, there is correlation between each special quality and its performance.

**Principal component analysis of the special qualities**

The seven special qualities, 100m run, standing triple jump, the run-up hand tough height, 4-6 strides approach height, after throwing the shot, the high grip barbell and squat barbell are expressed as $x_1, x_2, \cdots, x_7$ respectively. Conduct principal component analysis of the seven variables, and partial correlation coefficients among variables are shown in Table 4.

Conduct significance test of each partial correlation coefficient, and $P$ is less than 0.05, indicating that there are correlation between variables. Therefore it is necessary to conduct the principal component analysis.

Table 3: Correlation analysis results between the special quality and the specific performance

<table>
<thead>
<tr>
<th>Special performance /m</th>
<th>100m run /s</th>
<th>Standing triple jump /m</th>
<th>Run-up hand touch height /m</th>
<th>4-6 steps run-up height /m</th>
<th>After throwing shot /m</th>
<th>High grip barbell /m</th>
<th>Squat barbell /m</th>
</tr>
</thead>
<tbody>
<tr>
<td>r</td>
<td>-0.928</td>
<td>0.995</td>
<td>0.983</td>
<td>0.990</td>
<td>0.983</td>
<td>0.981</td>
<td>0.994</td>
</tr>
<tr>
<td>P</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Table 4: The statistical information of the principal component

<table>
<thead>
<tr>
<th>Component</th>
<th>Total</th>
<th>Initial Eigenvalues</th>
<th>Extraction Sums of Squared Loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>% of Variance</td>
<td>Cumulative %</td>
</tr>
<tr>
<td>1</td>
<td>6.745</td>
<td>96.360</td>
<td>96.360</td>
</tr>
<tr>
<td>2</td>
<td>0.154</td>
<td>2.202</td>
<td>98.562</td>
</tr>
<tr>
<td>3</td>
<td>0.069</td>
<td>0.981</td>
<td>99.543</td>
</tr>
<tr>
<td>4</td>
<td>0.018</td>
<td>0.252</td>
<td>99.795</td>
</tr>
<tr>
<td>5</td>
<td>0.009</td>
<td>0.124</td>
<td>99.919</td>
</tr>
<tr>
<td>6</td>
<td>0.004</td>
<td>0.058</td>
<td>99.978</td>
</tr>
<tr>
<td>7</td>
<td>0.002</td>
<td>0.022</td>
<td>100.000</td>
</tr>
</tbody>
</table>

The statistical information of the principal components is shown in Table 4. As can be seen from Table 4, the characteristic root of the first principal component is 96.36%, meaning the cumulative contribution rate reaches to 96.36%. Thus only select a principal component. The gravel figure of principal component analysis result is shown in Figure 1.

Matrix of factor loadings of the principal component analysis is shown in Table 5. As can be seen from Table 5, the first principal component contains the information of all the variables.

Judging from the matrix of factor loadings, expressions of the first principal component can be:

$$z_1 = -0.932x_1 + 0.994x_2 + 0.992x_3 + 0.984x_4 + 0.982x_5 + 0.988x_6 + 0.996x_7.$$

According to the above equation, the value of the first principal component can be obtained.
The first principal component as independent variables and high jumpers’ special performance as dependent variables, conduct the regression analysis. The coefficient of the model is $R^2 = 0.995$, indicating good fitting effect. The parameters of the model are shown in TABLE 6. Judging from TABLE 6, the final established regression equation is:

$$y = 2.335 + 0.042x_1 + 0.932x_2 + 0.994x_3 + 0.984x_4 + 0.992x_5 + 0.984x_6 + 0.988x_7 + 0.996x_8$$

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Partial regression coefficient</th>
<th>Standardized partial Regression coefficient</th>
<th>$t$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>2.335</td>
<td>-</td>
<td>2923.290</td>
<td>.000</td>
</tr>
<tr>
<td>$b$</td>
<td>0.042</td>
<td>0.998</td>
<td>50.348</td>
<td>.000</td>
</tr>
</tbody>
</table>

**CONCLUSIONS**

(1) In this study, the 100m run, standing triple jump, the run-up hand tough height, 4-6 strides approach high, after throwing shot, high grip barbell and squat barbell are closely related with the high Jumpers’ special performance.

(2) Build the prediction model of elite high Jump athletes:

$$y = a + bx$$

, respectively represents for the 100m run, standing triple jump, the run-up hand tough height, 4-6 stride approach height, after throwing shot, high grip barbell and barbell squat. This model reflects the relationship between high jump athletes’ special performances and special qualities and provides certain reference for the day-to-day training of the athletes.

**REFERENCES**


