International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 5080-5091

https://doi.org/10.70517/ijhsa464350

Assessment of technical gaps in high-level table tennis matches based on a multiple regression model: A case study

Xia Chen[1], Huagen Yin[2,*], Yanxiang Zhou[3] and Lin Zhou[4]
[1] School of Physical Education, Putian University, Putian, Fujian, 351100, China
[2] College of Physical Education, Shangrao Normal University, Shangrao, Jiangxi, 334001, China
[3] Shangrao Health Vocational College, Shangrao, Jiangxi, 334600, China
[4] East University of Heilongjiang, Harbin, Heilongjiang, 610043, China
Corresponding authors: (e-mail: yinhuagen1990@163.com).

Abstract The competitive level of modern table tennis is constantly improving, and the scientific assessment of athletes' technical and tactical abilities has become a key link to improve the training effect and game performance. In this paper, a multivariate statistical analysis method was used to construct a comprehensive strength assessment model for the technical and tactical abilities of high-level table tennis players. The study selected 155 singles matches of the world's top 20 male and female offensive players as the research samples, extracted the four main tactical factors of holding ability, serving scoring efficiency, stability and breaking ability by using factor analysis, and established a multiple linear regression model to assess the probability of winning. The results of the study showed that the KMO value was 0.698, the cumulative variance contribution of the four male factors amounted to 66.439%, and the coefficient of determination of the multiple regression equation, R^2, was 87.1%, which indicated that the tactical factors could explain 87.1% of the reasons for the changes in the probability of winning. Pass-through analysis revealed that there was an interaction effect among the tactical factors, and the direct pass-through coefficients were ranked as serve scoring efficiency (0.523), holding ability (0.522), stability (0.427), and breaking ability (0.246). The model validation showed that the overall technical and tactical indicators of men's singles were better than those of women's singles, and the scores of the holding ability factor were 0.652 and 0.759, respectively. The assessment model constructed in this study can objectively quantify the differences in the technical and tactical levels of the athletes and provide data support for the scientific training and game strategy development.

Index Terms Multiple regression model, factor analysis, technical and tactical assessment, through-trail analysis, table tennis, probability of winning

I. Introduction

Table tennis is a skill-driven, net-combat sport, known as the “national sport” in China, with a broad popular base and far-reaching cultural heritage. The reason why China's table tennis program has been able to maintain a long-lasting position in the international table tennis arena, in addition to the excellent cultural tradition of table tennis, the key lies in the emphasis on scientific research, scientific training through the support of science and technology, which plays an important role in guaranteeing the high efficiency of the national table tennis team to participate in the tournament and to create a good performance [1]-[4].

Currently, the focus of research on table tennis game is mainly concentrated in the field of technique and tactics, because technique and tactics are the dominant factors in the performance of competitive ability of table tennis players [5], [6]. Meanwhile, with the rapid development of modern information technology, exploring the winning rules of competitive sports programs and the construction of technical evaluation system through the cross integration of multiple disciplines has become the mainstream of research in the field of technical and tactical [7]-[9]. As the technical and tactical analysis and evaluation of oneself or one's opponent have always been an important source of information for table tennis players and coaches to design pre-match targeted training, on-field instruction, and to conduct other related scientific research [10]-[12]. Therefore, a table tennis player evaluation model can be constructed to reflect the individual characteristics of athletes in terms of technical and tactical aspects, physical fitness, psychology, and pre-match preparation, and then compare the strength differences between different athletes [13], [14].

In this study, a comprehensive technical and tactical strength assessment model was constructed by using multivariate statistical analysis, based on the match data of the world's high-level table tennis players. Firstly, factor analysis was applied to extract the main factors from many technical and tactical indicators to reduce the dimensionality of the data and reveal the underlying structure. Secondly, a multiple linear regression model was established to quantify the influence of each tactical factor on the probability of winning, which provided a mathematical basis for the strength assessment. The direct and indirect relationships between tactical factors are analyzed by pass-through analysis to reveal the interactive effects of tactical elements. Finally, the validity and practicability of the model is verified by using actual game data to ensure the reliability and application value of the research results.

II. Research design

II. A. Subjects of study

In this paper, a total of 155 singles matches (based on the world rankings published by the FIVB in March 2024) of the top 20 men's and top 20 women's world rankings of offensive playing athletes were selected. The match data were divided into two parts, the first part of 122 matches dated in 2022-March 2024, of which 10 matches (5 women's singles and 5 men's singles) were used for factor analysis and prediction test of multiple regression model, and the remaining 112 matches (52 men's singles and 60 women's singles) were used for the division of evaluation criteria of the three-dimensional model and the establishment of multiple regression model. The second part of the data, with match dates from early 2024 to early 2025, consisted of 33 matches (2 matches were duplicated with the first part of the data), including 20 men's singles and 13 women's singles, which were used to explore the impact of the implementation of the new rules on the technical and tactical utilization of excellent table tennis players.

II. B. Research methodology

II. B. 1) Video observation method

All the game videos in this study were downloaded from the official website of the International Table Tennis Federation (ITTF) http://www.ittf.com and the official website of CCTV5 http://cctv.cntv.cn the video viewing method was used to observe the points of the last beat of one athlete in each round, and the corresponding table was compiled by Excel for the original statistical record. Excel is used to calculate and count various indicators, and the obtained data is imported into SPSS18.0 for the verification of missing values, outliers and consistency.

II. B. 2) Mathematical modeling approach

Mathematical modeling method was used to establish the factor analysis and multiple linear regression model of tactical ability of excellent table tennis players, aiming at exploring the interrelationships among the indicators of the three-dimensional assessment model and the analysis of match simulation with different combinations of technical and tactical levels.

(1) Data sources

Fifty-two men's singles matches and 60 women's singles matches were selected from the database of this study for modeling, and another five women's singles and five men's singles matches were selected as new samples for model testing.

(2) Indicators and their calculation methods

“Scoring rate” is a qualitative characteristic of the technical and tactical nature of the table tennis game, and it is the optimal indicator to reflect the ability of the tactical factors, X1 (serve steal scoring rate), X2 (receive steal scoring rate), X3 (serve-round hold scoring rate, i.e., Holding I scoring rate), and X4 (receive-serve-round hold scoring rate, i.e., Holding II scoring rate) are taken as the tactical factors of the four indicators. The outcome of the game is measured by Y (probability of winning) as the dependent variable and output level indicator. The probability of winning is the percentage of the field score to the total number of points scored and lost in that game.

(3) Factor Analysis

Factor analysis is a multivariate statistical method commonly used for data downscaling and structural analysis, which describes the structure of the data by converting a set of highly correlated variables into a few irrelevant or weakly correlated factors. In the process of factor analysis, it is necessary to standardize the data, then apply mathematical methods such as principal component analysis to determine the minimum number of factors that can explain the variance of the data, and correlate the original variables with the factors to determine the corresponding factor loadings (i.e., correlation coefficients between the factors and the variables) for each variable. The factor analysis method can find the hidden representative factors among many variables and avoid the correlation between the sample data indicators from influencing the analysis results [15], and the basic model is as follows:

$$\begin{cases} X_1 = a_{11}F_1 + a_{12}F_2 + \cdots a_{1p}F_p + \varepsilon_1 \\ X_2 = a_{21}F_1 + a_{22}F_2 + \cdots a_{2p}F_p + \varepsilon_2 \\ \cdots \quad \cdots \\ X_n = a_{n1}F_1 + a_{n2}F_2 + \cdots a_{np}F_p + \varepsilon_n \end{cases} \tag{1}$$

Let there be an indicator marker quantity $X = (X_1, X_2, X_3, ..., X_n)$, the transformed common factor variables $F = (F_1, F_2, F_3, ..., F_p)$ (where $p < n$), with the value mean $E(F) = 0$, indicating no correlation between the components. E denotes the special factor of X, the common and special factors are independent of each other, and the covariance is denoted as $Cov(F, E) = 0$. $A(a_{ij})$ is the final factor loading matrix.

(4) Multiple linear regression model

The study of the relationship between two or more independent variables and a dependent variable constituting a quantitative change under the condition of linear correlation is called multiple linear regression analysis. Expression of this quantitative relationship of mathematical formulas, known as multiple linear regression model, its calculation usually need to be completed with the help of computers.

Let the linear regression model of random variable y and general variables $x_1, x_2, \cdots x_k$ be:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_k x_k + \varepsilon \tag{2}$$

where $\beta_0, \beta_1, \cdots, \beta_k$ are $k+1$ unknown parameters, β_0 is called the regression constant, and $\beta_1, \cdots, \beta_k$ is called the regression coefficient. y is called the explanatory (dependent) variable. The $x_1, x_2, \cdots, x_k$ are k general variables that can be measured and controlled and are called explanatory variables (independent variables). The regression coefficients in the multiple linear regression model are estimated using the least squares method. From the residual sum of squares $SSE = \sum (y - \hat{y})^2$, it can be seen that there is a minimal value of the residual sum of squares, SSE, according to the principle of finding the minimal value in calculus. To minimize SSE, the partial derivatives of SSE with respect to $\beta_0, \beta_1, \ldots, \beta_k$ must be zero. The partial derivatives of $\beta_0, \beta_1, \ldots, \beta_k$ are obtained by taking the SSE and making it zero, and organizing the partial derivatives into $k+1$ equations, which are then solved to obtain $\beta_0, \beta_1, \ldots, \beta_k$ for the estimates $\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_k$. Since the calculation is very complicated, statistical software is utilized to find the calculation results.

The construction of the multiple linear regression model requires testing the overall equation for goodness of fit, testing the overall regression equation and the significance of the regression coefficients.

(1) Goodness-of-fit test

For the degree of fit of multiple linear regression, the coefficient of multiple determination is used. It is defined as:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = 1 - \frac{\sum (y - \hat{y})^2}{\sum (y - \overline{y})^2} \tag{3}$$

where SSR is the sum of squares of regression, SSE is the sum of squares of residuals, and SST is the sum of squares of total deviations. $0 \le R^2 \le 1$, the closer R^2 is to 1, the higher the regression plane fit, and vice versa. In the actual regression analysis, an increase in the number of independent variables can make the regression sum of squares SSR increase, making R^2 increase, but this has nothing to do with the goodness of fit. Therefore, comparing the degree of fit between regression equations with different numbers of independent variables requires a correction or adjustment to R^2. The numerator and denominator of the ratio of the sum of squares of residuals to the sum of squares of total deviations are divided by their respective degrees of freedom to become the ratio of mean squared deviations to remove the effect of the number of independent variables on the degree of fit, which is adjusted to be $\overline{R}^2$, and the larger $\overline{R}^2$ is, the better it is in the realistic regression analysis.

(2) Significance test of overall regression equation

The significance test of the multiple linear overall regression equation is generally determined by the F test, using analysis of variance (ANOVA). The F statistic is defined as the ratio of the mean regression square to the mean sum of squares of the residuals (the mean square error), for a multiple linear regression equation:

$$F = \frac{SSR / k}{SSE / (n - k - 1)} \tag{4}$$

where n is the number of samples and k is the number of independent variables. The F statistic obeys an F distribution with k first degree of freedom and $n-k-1$ second degree of freedom. If the value of F is larger, it means that the change in the dependent variable caused by the independent variables is much larger than the effect of random factors on the formation of the dependent variable. Another way to see this is that the better the fit of the overall regression equation, the more significant the F statistic is. If $F \succ F_a(k, n-k-1)$ (or $p \prec \alpha$), the original hypothesis is rejected, indicating that there is a significant linear relationship between the independent variable and the dependent variable, and the overall regression equation is significant.

(3) Significance test of regression coefficients

The purpose of the significance test of regression coefficients is to verify that the respective variables $x_1, x_2, ..., x_k$ on the dependent variable y is significant or not by using t test. For multiple linear regression equations:

$$t = \frac{\beta_i}{S_{\beta_i}} = r(n-k-1) \tag{5}$$

where n is the number of samples, $n-k-1$ is the degrees of freedom, and S_{β_i} is the standard error of the regression coefficient β_i. If the standard error of the regression coefficient of an independent variable is large, then a relatively small value of t is inevitably obtained, indicating that the independent variable is poorly able to explain the change in the dependent variable. Therefore, when the t value of an independent variable is small enough, that independent variable should not be retained. Given the level of significance α, determine the critical value $t_a(n-k-1)$ or calculate the size of the concomitant probability value p that corresponds to the t value. If $|t| \succ t_{\bar{\alpha}}(n-k-1)$ (or $p \prec \alpha$), the regression coefficient is considered significantly different from zero, i.e., there is a significant relationship between this independent variable and the dependent variable.

Finally, the underlying assumptions and outliers were tested using SPSS 17.0 statistical software.

II. B. 3) Methods of statistical analysis

A series of statistical analyses of the relevant data indicators of this study were carried out using SPSS and DPS statistical software:

(1) Descriptive statistics: the summary and organization of the whole set of data involved in the study is generally expressed as mean ± standard deviation.

(2) Percentile method: the percentile method was used to develop the evaluation criteria of the model indicators. The data of 52 men's singles matches and 60 women's singles matches were initially counted, and then the percentile method was used to divide each assessment indicator into different proportions.

(3) Chart-square test: the significance test of the two sets of percentile data (scoring rate, utilization rate, usage degree, and contribution degree) was used.

(4) ANOVA: The comparison of technical efficiency indicators was analyzed by ANOVA, and all data were tested to be in line with normal distribution.

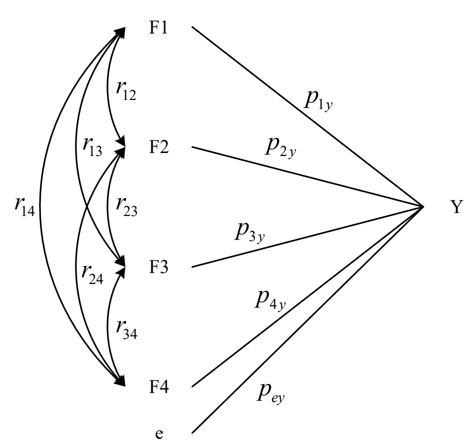

Figure 1: Table tennis tactical factor path analysis model

(5) t-test: Independent sample t-test was used for the comparative analysis of the strength difference indicators of the two groups, and paired sample t-test was used for the technical benefit indicators of the excellent male and female athletes before and after the use of the new plastic ball. After the test, the overall sample of each technical benefit index obeyed normal distribution.

(6) Pass-through analysis: Pass-through analysis is based on multiple linear regression analysis, through further analysis and decomposition of standardized regression coefficients, to reveal the direct and indirect effects of the

main factors affecting the dependent variable, and to provide scientific and reasonable quantitative explanations for the ways and means of the role of the respective variables [16]. The relationship between the tactical factors (i.e., the four-segment scoring rate, F1, F2, F3, F4) and the probability of winning Y in this study was analyzed by the through-path analysis to reveal more interactive features between the indicators. The model for establishing the through-path analysis is shown in Figure 1, where one-way arrows indicate the causal relationship between the independent or latent variable and the dependent variable, two-way arrows indicate the correlation between two independent variables, and e indicates the direct effect of random error on the dependent variable Y.

II. B. 4) Two-party adversarial dimension algorithm

Based on the correspondence between the technical scores and losses of both players in a table tennis match, the concept of “strength difference” is proposed, which is essentially the difference between the percentage of one player's segment (or round, or court) scores in the total number of points scored or lost in the whole match, and the percentage of the opponent's scores in the corresponding segment (or round, or court) in the total number of points scored or lost in the whole match, and is assessed as the difference in strength between the two players during the confrontation. The assessment is the strength difference between the two athletes in the confrontation process, so it can also be called “Strength Difference Assessment Method”.

(1) Calculation method of each index of A-side athletes

Let I be the sum of the total points scored and total points lost in the game, i.e. $I = A + B + C + D + X + Y + Z$ (the same as the following), then the formula of the strength difference of A-side athletes is expressed as follows:

a) Calculation of strength difference of four segments

Hair grabbing segment strength difference:

$$(SD_{1-3}) = \frac{(A^+ + B^+) - (A^- + B^- + C^-)}{I} \times 100\% \tag{6}$$

Poor strength in catch-and-snatch segments:

$$(SD_{2-4}) = \frac{(X^+ + Y^+) - (X^- + Y^-)}{I} \times 100\% \tag{7}$$

Holding I segment strength is poor:

$$(SD_{z5}) = \frac{(C^+ + D^+) - D^-}{I} \times 100\% \tag{8}$$

Holding II segment strength difference:

$$(SD_{\geq 6}) = \frac{Z^+ - Z^-}{I} \times 100\% \tag{9}$$

b) Calculation of the strength difference of the three segments

Strength difference of hair grab segment = (SD_{1-3}), same as formula (6).

Receiving snatch segment strength difference = (SD_{2-4}), same as formula (7).

Holding segment strength difference:

$$(SD_{\geq 5}) + (SD_{\geq 6}) \tag{10}$$

c) Calculation of wheel strength difference

Serve round strength difference:

$$(SD_{1-3}) + (SD_{\geq 5}) \tag{11}$$

Poor receiving wheel strength:

$$(SD_{2-4}) + (SD_{\geq 6}) \tag{12}$$

d) Calculation of the total strength difference:

$$(SD_{1-3}) + (SD_{2-4}) + (SD_{\geq 5}) + (SD_{\geq 6}) \tag{13}$$

(2) Calculation method for each index of side B athletes

According to the correspondence between the athletes of both sides in the match, the difference in the strength of each segment of the athletes of side B is the opposite of the difference in the strength of the corresponding batting segment of the athletes of side A.

II. B. 5) Overall strength dimension algorithm

The overall strength dimension adopts the three-section structure of the start-stealing section, the receiving section and the holding section, and adopts the two indicators of the degree of use and the degree of contribution to emphasize the proportion distribution of the athlete's technical use and scoring in the three sections of the whole game (or multiple comprehensive games). The degree of utilization is the percentage of the total points scored and lost in each segment to the total points scored and lost in the whole game, and the degree of contribution is the percentage of the total points scored in each segment to the total points scored in the whole game, and the indexes of the degree of utilization and the degree of contribution of each segment are calculated as follows:

(1) Calculation method of each index for A-side athletes:

The degree of use of the A-side serve-and-snatch segment:

$$\frac{A+B+C^-}{I}\times 100\% \tag{14}$$

Party A picks up the degree of robotic segment usage:

$$\frac{X+Y}{I}\times 100\% \tag{15}$$

The degree of use of the A-side holding segment:

$$\frac{C^+ + D + Z}{I}\times 100\% \tag{16}$$

A-square hair grabbing segment contribution:

$$\frac{A^+ + B^+}{I^+}\times 100\% \tag{17}$$

Side A picks up the snatch segment contribution:

$$\frac{X^+ + Y^+}{I^+}\times 100\% \tag{18}$$

Contribution of A-side holding segments:

$$\frac{C^+ + D^+ + Z^+}{I^+}\times 100\% \tag{19}$$

(2) Calculation method of each indicator for Side B athletes

The value of each index of B-square athletes can be calculated according to the data of A-square athletes:

The degree of use of the hair-snatch segment of side B = the degree of use of the catch-snatch segment of side A:

$$\frac{X+Y}{I}\times 100\% \tag{20}$$

The degree of use of the receiving segment of party B = the degree of use of the sending segment of party A:

$$\frac{A+B+C^-}{I}\times 100\% \tag{21}$$

B-side holding segment utilization = A-side holding segment utilization:

$$\frac{C^+ + D + Z}{I}\times 100\% \tag{22}$$

B-side hair snatch segment contribution:

$$\frac{X^{-}+Y^{-}}{I^{-}}\times 100\% \tag{23}$$

B-side pickup snatch segment contribution:

$$\frac{A^{-}+B^{-}+C^{-}}{I^{-}}\times 100\% \tag{24}$$

B-side phase-in contribution:

$$\frac{D^{-}+Z^{-}}{I^{-}}\times 100\% \tag{25}$$

III. Findings and analysis

III. A. Comprehensive Strength Evaluation of Table Tennis Singles

III. A. 1) Correlation test between singles techniques

The KMO and Bartlett sphericity tests are shown in Table 1. It can be seen that the correlation test of each technical and tactical indicators was tested using KMO and Bartlett sphericity test. The value of KMO was measured to be 0.698, which indicates that this data is suitable for factor analysis because it is closer to 0.7 and because P=0.00<0.05.

Table 1: KMO and Bartlett sphericity tests

KMO sampling availability number	Bartlett's Test of Sphericity		
0.698	Chi-square	F=53	P=0.000

III. A. 2) Technical and Tactical Factor Analysis for Table Tennis Singles

In accordance with the principle of extracting eigenvalues greater than 1, the principal component analysis method inside the factor analysis was used to extract a total of four common factors, and the interpretation of the variance of each party after rotation is shown in Table 2. The variance contribution rates of common factors 1 to 4 after rotation were 20.004%, 18.081%, 14.672% and 13.682%, and the eigenvalues of the four common factors were 2.221, 1.979, 1.608 and 1.489, respectively.

Table 2: The explanation of the difference in the rotation

	Rotational load squared		
	Total	Percentage of variance	Cumulative (%)
1	2.221	20.004	20.008
2	1.979	18.081	38.08
3	1.608	14.672	52.759
4	1.489	13.682	66.439

Orthogonal rotation of the initial component matrix using the maximum variance method allowed for a deeper understanding and interpretation of the actual significance of the four common factors. The rotated component matrix of the game data is shown in Table 3. Through the data in Table 3, we know the degree of relationship between each technical and tactical index and the common factors, and the larger the correlation coefficient, the higher the correlation between them. Through the analysis of the rotated factor component matrix, in the common factor 1, according to the index loading coefficient from the largest to the smallest are running distance, unforced errors, winning points. The fewer unforced turnovers proved that the stronger the holding and running ability, so it was named the holding ability factor. The indicator loading coefficients of the public factor 2 in descending order are first serve scoring rate, Aces, second serve scoring rate, through the analysis of these three indicators are determined by the actual scoring rate of the player's serve, so we named the public factor 2 as the efficiency of serve scoring factor. The load factor of each variable indicator of the public factor 3 from small to large is the success rate of the first serve, double faults, over the line scoring rate, named as stability factor. Common factor 4 from small to large are the success rate of break serve, the rate of receiving serve, the rate of receiving serve is closely related to the success rate of break serve, so the two technical and tactical indicators are named as the break serve ability factor.

Table 3: The game data rotation composition matrix

Original index variable	Constituent			
	1	2	3	4
ACES (X1)	0.329	0.693	-0.297	-0.136
Double error (X2)	0.272	0.045	-0.728	0.21
Success rate (X3)	0.083	-0.086	0.78	0.108
One point (X4)	-0.089	0.863	-0.024	0.137
Second rate (X5)	-0.26	0.541	0.282	0.138
Access rate (X6)	0.067	0.233	0.401	0.361
Break rate (X7)	-0.015	-0.094	-0.175	0.809
Receiving rate (X8)	0.009	0.199	0.161	0.778
Winning (X9)	0.749	0.509	-0.055	0.006
Involuntary error (X10)	0.772	-0.257	-0.268	-0.049
Running distance (X11)	0.885	-0.014	0.123	0.061

III. A. 3) Comprehensive singles ability factor modeling

The combined singles characterization factor score coefficients are shown in Table 4. Based on the data in the following table the model formula for each factor score can be derived as:

$$F1 = 0.094X1 + 0.051X2 + 0.151X3 + \cdots + 0.437X11 \quad (26)$$

$$F2 = 0.367X1 - 0.073X2 - 0.07X3 + \cdots - 0.065X11 \quad (27)$$

$$F3 = -0.152X1 - 0.455X2 + 0.527X3 + \cdots + 0.214X11 \quad (28)$$

$$F4 = -0.138X1 + 0.172X2 + 0.006X3 + \cdots + 0.017X11 \quad (29)$$

$$F_{Total} = (0.24F1 + 0.1763F2 + 0.1436F3 + 0.1365F4) / 0.6523 \quad (30)$$

According to the comprehensive score factor model can be more accurately evaluated the comprehensive score ability of each player.

Table 4: Integrated single-beat feature factor score coefficient

Original index variable	Constituent			
	1	2	3	4
ACES (X1)	0.094	0.367	-0.152	-0.138
Double error (X2)	0.051	-0.073	-0.455	0.172
Success rate (X3)	0.151	-0.07	0.527	0.006
One point (X4)	-0.099	0.441	-0.052	0.009
Second rate (X5)	-0.108	0.28	0.168	0.027
Access rate (X6)	0.061	0.093	0.241	0.186
Break rate (X7)	-0.034	-0.129	-0.174	0.586
Receiving rate (X8)	0.006	0.028	0.053	0.505
Winning (X9)	0.311	0.216	0.042	-0.057
Involuntary error (X10)	0.349	-0.172	-0.072	0.012
Running distance (X11)	0.437	-0.065	0.214	0.017

III. B. Multiple regression analysis of tactical ability

Previous studies have shown that the scoring rate of table tennis players in each competitive interval segment of a match directly affects or indirectly affects the probability of winning the match, therefore, this study decided to construct a multiple regression equation by setting the independent variables in the multiple regression analysis to be the ability to hold (F1), the efficiency of the serve scoring (F2), the consistency (F3), and the ability to break the serve (F4), and the dependent variable to be the probability of winning the match (Y) The dependent variable is the probability of winning (Y), thus constructing a multiple regression equation. This multiple regression equation can not only compare the importance of tactical ability to the probability of winning in each competitive interval, but also construct a model of tactical ability that is suitable for the best combination of table tennis singles matches.

III. B. 1) Construction of multiple regression equations

F1, F2, F3 and F4 were set as independent variables and Y was set as the dependent variable, and the data were entered into the SPSS software by selecting the method "Enter". The calculation results are shown in Tables 5 and 6.

Table 5: Multiple Regression Model

Model	R	R^2		Adj_ R^2	SD	F	Durbin-Watson
Singles	0.963[a]	0.871		0.866	0.021	46.326	1.636

Table 6: Model Coefficient Test Results

Model	Unnormalized coefficient		Normalization factor		Sig	Common linear diameter	
	B	SD	Beta	T		Tolerance	VIF
Constants	0.013	0.055	-	0.383	0.714	-	-
F1	0.29	0.038	0.522	7.175	-0.015	0.944	1.08
F2	0.394	0.056	0.523	7.378	0.004	0.937	1.061
F3	0.238	0.024	0.427	5.925	-0.006	0.882	1.144
F4	0.098	0.026	0.246	3.613	0.01	0.939	1.083

III. B. 2) Tests of multiple regression equations

Before and after performing the multiple regression analysis, the data need to be examined to see if they satisfy the 7 tests.The 7 tests are (1) The dependent variable is a continuous variable. (2) There are no less than 2 independent variables. (3) There are mutually independent observations. (4) There is a linear relationship between the independent and dependent variables. (5) Equal variance. (6) There is no multicollinearity. (7) Residuals are approximately normally distributed.

(1) The dependent variable of the multiple regression equation in this paper is a continuous variable and the equation contains 4 independent variables, thus meeting the first 2 tests.

(2) The Durbin-Watson test value requirement is generally 0 to 4, and the closer the value is to 2, the greater the possibility that the observations are independent of each other. Based on the results, it can be seen that the Durbin-Waston value in this study is 1.664, which indicates that the observations are more independent from each other, thus meeting the test 3.

(3) The correlation test was conducted on the four independent and dependent variables in the equation and the results are shown in Table 7. The Pearson correlation coefficients between the independent variables and the dependent variable are 0.426, 0.639, 0.452 and 0.455, respectively, indicating that there is a linear correlation between the four independent variables and the dependent variable, which is in line with test 4.

Table 7: Pearson Coefficients

	F1	F2	F3	F4	Y
F1	1	0.036	-0.236	-0.059	0.426*
F2	0.036	1	0.163	0.166	0.639**
F3	-0.236	0.163	1	0.186	0.452*
F4	-0.059	0.163	0.186	1	0.455*
Y	0.426*	0.639**	0.452*	0.455*	-

(4) Judging from the value of f in the equation, the equation is isotropic, which conforms to test 5.

(5) The expansion factors (VIF) in the multivariate equations are less than 2, which proves that there is no multicollinearity relationship in the equations and conforms to test 6.

(6) The residuals of the regression equations in this paper are tested for normality using histograms, which is the easiest and most intuitive method in the normality test of regression equations. The standardized residuals of the regression constructed in this study are shown in Figure 2. It can be seen that the residuals are approximately normally distributed and therefore meet the test 7.

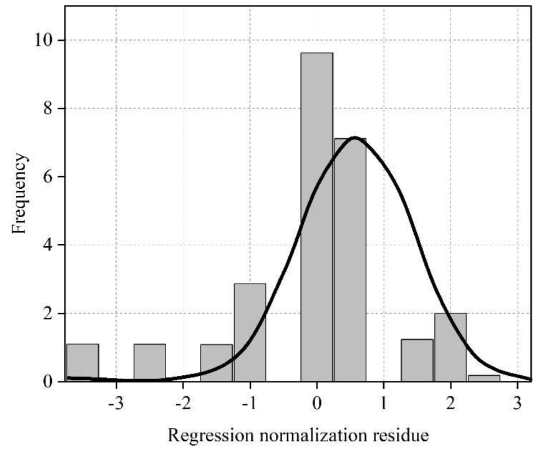

Figure 2: Multiple Regression Residuals

This equation passes the above seven tests, indicating that the calculations of this multiple regression analysis are reasonable, meaningful and credible.

III. B. 3) Calculation of multiple regression analysis

In multiple regression, the coefficient of determination R^2 indicates the degree of fit of the equation, i.e.: the extent to which changes in the dependent variable can be explained by the regression equation, according to Table 5, it can be seen that the coefficient of determination of the regression equation established in this study is 87.1%, and this data suggests that the four independent variables explain 87.1% of the reasons for the changes in the dependent variable, i.e.: the performance of the tactical ability of the four competitive interval segments of the game explains 87.1% of the reasons for the changes in the probability of victory 87.1% of the reasons for the change in size. This indicates that the equation has a good fit.

III. B. 4) Multiple regression modeling

Table 6 shows the results of the calculation of the tactical ability model for the competitive interval segment of the competition, and the significance results of the respective variables in the equation are less than or equal to 0.01, which shows that the equation is statistically significant and the equation passes the various tests, which proves that the equation is valid. Therefore, it can be concluded that the competitive interval segment tactical ability model is based on the calculation results:

$$Y = 0.013 + 0.29F1 + 0.394F2 + 0.238F3 + 0.098F4 \tag{31}$$

III. C. Calculation of indirect flux coefficients

In statistics, the indirect path coefficient indicates the effect of one independent variable on the dependent variable through other independent variables. Therefore, this paper will introduce this concept to explore the influence of the tactical ability of each competitive interval on the probability of winning the game through the competitive ability of the rest of the competitive segments, and whether the tactical ability of each competitive interval affects each other. The theory of path analysis has proved that the indirect path coefficient of any independent variable Fi to Y is equal to the product of the correlation coefficient and the path coefficient. As can be seen from Table 6, the normalized coefficients in the regression equation are P1y=0.522, P2y=0.523, P3y=0.427, and P4y=0.246, respectively. From Table 7, it can be seen that the correlation coefficients between the respective variables are r12=R²1=0.036, r13=r31=-0.236, r14=r41=-0.059, R²3=r32=0.426, R²4=r42=0.163, r34=r43=0.186, respectively. According to the formula, the indirect coefficient of F1 to Y generated by F2 is equal to r12*P2y=0.036*0.523=0.018828. The indirect coefficient of F1 to Y generated by F3 is equal to r13*P3y=-0.236*0.427=-0.100772. The indirect coefficient of F1 to Y generated by F4 is equal to -0.059*0.246=-0.014514. In the same way, the indirect diameter coefficients of F2, F3, and F4 for Y can be obtained from this formula. The calculation results are shown in Table 8. The order of the direct path coefficient of the four independent variables is inconsistent with the order of the correlation coefficient, the order of the direct path coefficient is F2>F1> F3>F4, and in the ranking of the correlation coefficient, F2>F3> F1>F4, the ranking of the grabbing segment drops to the third place, which indicates that the indirect effect has a certain impact in the singles game.

Table 8: Indirect Path Coefficient Values

Independent variable	Correlation coefficient of Y	Path coefficient	Indirect path coefficient				
			F1	F2	F3	F4	Total
F1	0.411	0.526	-	0.018828	-0.100772	-0.014514	-0.09646
F2	0.653	0.522	0.021	-	0.086	0.043	0.15
F3	0.452	0.432	-0.112366	0.08635	-	0.050636	0.02462
F4	0.401	0.251	-0.036985	0.09635	0.086363	-	0.14573

IV. Application Analysis of Table Tennis Technical and Tactical Evaluation Models

The model constructed above will be used to analyze and evaluate table tennis singles matches. The athletes on both sides of the matches selected for this study are Chinese high-level table tennis players, so they have good research value. Two athletes who participated in men's singles and women's singles were selected for the national matches in different years, which were the singles finals of the 2018 National Championships and the singles finals of the 2022 National Championships, and each of the 2 matches won 1 match, which is very suitable to be used as a case for model validation.

The observation results of scoring efficiency indicators are shown in Table 9. In summary, in the mixed doubles final of the 2022 National Championships, the men's singles is much higher than the women's singles from the overall view of the technical and tactical evaluation indicators, both in terms of the number of excellent indicators and the comparison of the indicators between the two sides. In addition, observing the scoring efficiency indicators of this match, although the men's singles is much larger than the women's singles, the technical and tactical use of the indicators of holding ability is still insufficient.

Table 9: Index observation results

Index	M	SD	Men's singles	Abnormality	Women's singles	Abnormality
F1	0.474	0.989	0.652	0	0.759	0
F2	0.552	0.868	0.898	0	0.816	0
F3	0.54	0.782	0.729	0	0.646	0
F4	0.446	0.774	0.772	0	0.636	0

In conclusion, the table tennis technical and tactical assessment model constructed in this paper has high practicality and accuracy.

V. Conclusion

The study successfully constructed a comprehensive strength assessment model of technique and tactics based on multiple regression by analyzing 155 high-level table tennis singles matches in depth.

The results of factor analysis showed that four main tactical factors, namely, holding ability, serving scoring efficiency, stability and breaking ability, were successfully extracted from 11 original tactical indicators, with a cumulative variance contribution rate of 66.439%, which effectively summarized the core tactical and technical characteristics of the athletes. The multiple regression analysis showed that the established prediction model fit well with a coefficient of determination of 87.1%, which passed the strict statistical test and proved that the model had high scientific validity and reliability. The pass-through analysis further revealed the complex relationship between the tactical factors, and the direct pass-through coefficient of serve-scoring efficiency was the highest (0.523), followed by phasing ability (0.522), indicating that these two factors had the most significant direct effect on the probability of winning. Indirect effect analysis revealed a significant interaction between the tactical factors, with the catch-and-snatch segment being most favorably affected by the interaction effect. The results of model validation showed that men's singles was outstanding in serving scoring efficiency (0.898), which was significantly better than women's singles (0.816), reflecting the different characteristics of athletes of different genders in the use of technology and tactics.

The evaluation model provides a quantitative tool for coaches to scientifically formulate training plans, optimize technical and tactical configurations, and accurately assess the strength level of athletes, which is of great theoretical value in promoting scientific training in table tennis.

Author contributions

Conceptualization, X C; methodology, HG Y; formal analysis, X C; data processing, X C and YX Z; writing—original draft preparation, X C; writing—review and editing, HG Y; project administration, X C; funding acquisition, N/A. All

authors have reviewed and agreed to the published version of the manuscript. All authors read and approved the final manuscript.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethical approval and consent to participate

Consent for publication

N/A.

Competing interests

The authors declare that they have no competing interests.

Funding

This work has been supported by the Economic and Social Development and Philosophy and Social Science Fund of Shangrao City (NO. 24YB23); the Education and Research Program for Young Teachers in Fujian Province (Science and Technology) (NO. JAT220308); and the Science and Technology Project Plan of Putian City (NO. 2023SZ3001PPTXY15); Putian University Research Project "Intervention and Mechanism Study of Fitness Yoga on Upper Cross Syndrome" (No. 2022056). The acquisition of experimental data from these competition videos complies with the current laws of the country. The author does not have a conflict of interest to declare.

References

[1] Wang, X. (2024, May). Research and practice on intelligent optimization model of table tennis tactics based on reinforcement learning algorithm. In International Conference on Artificial Intelligence for Society (pp. 447-456). Cham: Springer Nature Switzerland.

[2] Sung, C. Y. (2019). A smart analysis of technical skills of top male table tennis players. Smart Science, 7(4), 231-238.

[3] Yin, H., Zhou, Y., Chen, X., & Zhou, L. (2025). Construction and application of the technical and tactical efficiency evaluation model for table tennis matches based on grey correlation. BMC Sports Science, Medicine and Rehabilitation, 17(1), 1-14.

[4] Yu, X. (2023). Evaluation of training efficiency of table tennis players based on computer video processing technology. Optik, 273, 170404.

[5] Ivanek, V., Đukić, B., Mikić, B., Smajic, M., & Doder, D. (2018). Effects of technical and tactical characteristics on the performance of the table tennis players. Facta Universitatis, Series: Physical Education and Sport, 16(1), 157-166.

[6] Ganieva, M., Abduxalilova, S., Azimova, S., Yuldashaliyeva, M., & Kazakov, R. (2024). THE UNITY OF TECHNIQUE, TACTICS AND STRATEGY IN TABLE TENNIS. Modern Science and Research, 3(5), 1175-1183.

[7] Zhang, H., Zhou, Z., & Yang, Q. (2018). Match analyses of table tennis in China: a systematic review. Journal of sports sciences, 36(23), 2663-2674.

[8] Faber, I. R., Koopmann, T., Büsch, D., & Schorer, J. (2021). Developing a tool to assess technical skills in talented youth table tennis players—a multi-method approach combining professional and scientific literature and coaches' perspectives. Sports medicine-open, 7(1), 42.

[9] Fuchs, M., Liu, R., Malagoli Lanzoni, I., Munivrana, G., Straub, G., Tamaki, S., ... & Lames, M. (2018). Table tennis match analysis: a review. Journal of sports sciences, 36(23), 2653-2662.

[10] Yin, H., Chen, X., Zhou, Y., Xu, J., & Huang, D. (2023). Contribution quality evaluation of table tennis match by using TOPSIS-RSR method-an empirical study. BMC Sports Science, Medicine and Rehabilitation, 15(1), 132.

[11] Zhou, Z., & Zhang, H. (2022). A visible analysis approach for table tennis tactical benefit. Journal of Sports Science & Medicine, 21(4), 517.

[12] Zhang, S., Chen, G., Wu, Q., & Li, X. (2023). The interplay between table tennis skill development and sports performance: A Comprehensive review. Pacific International Journal, 6(3), 150-156.

[13] Pradas de la Fuente, F., Ortega-Zayas, M. Á., Toro-Román, V., & Moreno-Azze, A. (2023). Analysis of technical–tactical actions in high-level table tennis players: differences between sexes. Sports, 11(11), 225.

[14] Chen, Z., Li, M., & Yang, Q. (2025). Evaluation of elite table tennis players' comprehensive technical and tactical performance. BMC Sports Science, Medicine and Rehabilitation, 17(1), 45.

[15] Hashemi Farzane,Naderi Mehrdad,Jamalizadeh Ahad & Bekker Andriette. (2021). A flexible factor analysis based on the class of mean-mixture of normal distributions. Computational Statistics and Data Analysis,157,

[16] Mastaneh Mokayef,M.K.A Ahamed Khan,MHD Amen Summakieh,Lee Qi Jian & Miad Mokayef. (2023). Embedded Ball Launcher with Trajectory Path Analysis for Empowering Junior Table Tennis Players. Journal of Advances in Artificial Life Robotics,4(1),27-34.