ISSN : 0974 - 7435 Volume 8 Issue 9

BioTechnology
An Indian Journal

FULL PAPER

BTAIJ, 8(9), 2013 [1228-1233]

PLS aerobics training mathematical regression model applied research

Caiyan Gao[1], AbdolGhaffar Ebadi[2]
[1]JilinUniversity, Changchun 130012, Jilin, (CHINA)
1E-mail: tiyuxi@qq.com
[2]Department of Agriculture, Jouybar branch, Islamic Azad University, Jouybar (IRAN)

ABSTRACT

The spot to play in the game against athlete mistakes asked to explore the athletes in the race through the ways and means to complete their usual training a small mistake difficult action to achieve better race results. Establish Aerobics partial least squares regression modeling to improve forecasts on of aerobics athlete's intention and behavior. Studies have shown that exercise program, self-efficacy, to support the operation mechanism between intention and behavior, aerobics athlete's movement behavior intervention by increasing self-efficacy and support of motor behavior.

KEYWORDS

Aerobics;
Partial least Squares;
Regression model.

INTRODUCTION

Suppose in their daily practice, aerobics athletes realize the movement of the highest degree of difficulty can be Successfully completed with the probability of 50%, and at the crucial moment of competition, they will probably fail to finish this movement, how to face such potential threat? Many scholars have proposed some methods, such as adjusting breathing to regulate one's emotions, refraining oneself from thinking over the possibility of failure, and adjusting one's mental status to complete the movement smoothly, etc.

A great deal of research shows that the sole use of intention cannot predict the behavior change. This paper keeps on investigating the predictive effect of planning as a mediator variable. Self-efficacy theory is derived from Bandera's social-cognitive theory; it refers to individuals' beliefs and the expectations in their capabilities to perform a task successfully, or the sub consciousness of achieving success. The correlation between self-efficacy level and performance of aerobics athletes of the provincial level is 0.57, a previous study showed an average correlation of 0.38 between self-efficacy and performance according to path analysis. Athletes with prominent self-efficacy will exert differently in practice and competition, the same is true with aerobics athletes.

This paper measures the time, place and mode of respondents' practice based on the theory of planned behavior. To improve the accuracy of prediction, the model derived from the theory of planned behavior is modified by testing the effect of variables and adding other variables. This paper tests the validity of this model by partial least square regression, which has the advantage of estimating structural relationship with restricted error. Multifaceted latent variables of exercise self-efficacy and social support are the representation of the competition field. The defense mechanism of people aims at eliminating the self-integrity of threats. This paper tests the explanation and forecast effect of theoretical model with the initial model of theoretical planning behavior, and then train the model, taking self-efficacy and social support as variables introduced into partial least squares regression model, and the aim is to enhance the explanation and prediction of intention and behavior.

PARTICAL LEAST SQUARES REGRESSION MODEL

The rationale of curve-fitting problem is to find a most suitable function to match the relationship between independent variable (s) and dependent variable (s). In general, an established function expression with unknown factors is given according to prior knowledge or visual observation of data, and the main work is to get the factors. From the computational perspective, the problem seems to be completely solved, but further research is necessary. From the statistical perspective, all data are random variables, and the computation is based on sample data, so hypothesis testing is needed; if the interval estimation is too wide, or the result of factor is zero, the result will be meaningless. Moreover, error analysis of variance analysis model can be used to estimate the results. In short, regression is a statistical analysis method of fitting problem.

Partial least squares regression (PLS) is an advanced multivariable statistics analysis method, which is used to establish the correlations between multiple dependent variables. In a regression analysis problem containing many independent variables and dependent variables in which serious multicollinearity exists, results from general multiple regression method is low in reliability, and PLS regression is an ideal solution to this problem.

A modeling problem: p dependent variables: $y_1, y_2, ..., y_p$ and m independent variables: $x_1, x_2, ..., x_m$ is put into consideration. To analyze data by PLS regression, in the first step, the first principal component of independent variables t_1 which is a linear combination of $x_1, x_2, ..., x_m$ and contains as much information from original data as possible is extracted, and then the first principal component of dependent variables, u_1 is extracted which has the maximum correlation with t_1; in the next step, a regression between $y_1, y_2, ..., y_p$ and t_1 is built and the explanation degree is tested, if the explanation degree is precise enough, the computation is terminated, otherwise another principal component is extracted until the precision demand is satisfied. Finally, a PLS regression equation will be built between r components, $t_1, t_2, ..., t_r$ and dependent variables, $y_1, y_2, ..., y_p$.

For convenience of calculations, we suppose p dependent variables $y_1, y_2, ..., y_p$ and m independent variables $x_1, x_2, ..., x_m$ are all pre-standardized. From n observations, the values of variables are recorded in the following two matrixes:

$$F_0 = \begin{pmatrix} y_{11} & \cdots & y_{1p} \\ \vdots & \ddots & \vdots \\ y_{n1} & \cdots & y_{np} \end{pmatrix}, E_0 = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix}$$

The specific steps of PLS Regression analysis for the training of aerobics athletes are as follows:

Step 1: Extract the first principal components of independent variables and dependent variables respectively. Suppose they are t_1 and u_1. t_1 is the linear combination of independent variables $X = (x_1, ..., x_m)^T$: $t_1 = w_{11}x_1 + ... + w_{1m}x_m = w_1^T X$, and u_1 is the linear combination of independent variables $Y = (y_{1,\cdots,} y_p)^T$: $u_1 = v_{11}y_1 + ... + v_{1p}y_p = v_1^T Y 0$

As is required by PLS regression model, t_1 and u_1 must draw as much variation information as possible of the array in which they lie to maximize the correlation between them. The result vector of the first principal component pair, $\hat{t}_1$ and $\hat{u}_1$ can be determined by standardized observation data matrixes E_0 and F_0.

$$\hat{t}_1 = E_0 w_1 = \begin{pmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{mn} \end{pmatrix} \begin{pmatrix} w_{11} \\ \vdots \\ w_{1m} \end{pmatrix} = \begin{pmatrix} t_{11} \\ \vdots \\ t_{n1} \end{pmatrix},$$

$$\hat{u}_1 = F_0 V_1 = \begin{pmatrix} Y_{11} & \cdots & y_{1P} \\ \vdots & & \vdots \\ Y_{n1} & \cdots & y_{np} \end{pmatrix} \begin{pmatrix} v_{11} \\ \vdots \\ v_{1p} \end{pmatrix} = \begin{pmatrix} u_{11} \\ \vdots \\ u_{n1} \end{pmatrix}.$$

The covariance between the first principal

FULL PAPER

component pair is $Cov = (t_1, u_1)$ that is computed by the inner product of $\hat{t}_1$ and $\hat{u}_1$. Thus, this problem will be transformed into an extremum problem with subsidiary condition:

$$\begin{cases} \left\langle \hat{t}_1, \hat{u}_1 \right\rangle = \left\langle E_0 w_1, F_0 V_1 \right\rangle = w_1^T E_0^T F_0 x_1 \Rightarrow \max \\ w_1^T w = \|w_1\|^2 = 1, v_1^T v_1 = \|v_1\|^2 = 1 \end{cases}.$$

The above is to get the values of w_1 and v_1 when $\theta_1 = w_1^T E_0^T F_0 v_1 \Rightarrow$ maximizes. To solve this problem, the eigenvalue and eigenvector of the $m \times m$ matrix $M = E_0^T F_0 F_0^T E_0$ must be obtained in the first place. The greatest eigenvalue of M is θ_1^2 while the corresponding eigenvector equals to w_1, and v_1 can be obtained based on w_1 through the formula $v_1 = \frac{1}{\theta_1} F_0^T E_0 w_1$.

Step 2: The regression model between $y_1, ..., y_p$ and u_1 and that between $x_1, ..., x_m$ and t_1 must be constructed. The model is expressed as follows:

$$\begin{cases} E_0 = \hat{t}_1 \alpha_1^T + E_1 \\ F_0 = \hat{u}_1 \beta_1^T + F_1 \end{cases}$$

In the equations above, $\alpha_1 = (\alpha_{11}, ..., \alpha_{1m})^T$, $\beta_1 = (\beta_{11}, ..., \beta_{1p})^T$ are many-to-one parameter vectors which can be called effect loading vectors, while E_1 and F_1 are residual matrixes. The least squares estimation of α_1 and β_1 are as follows:

$$\begin{cases} \alpha_1 = E_0^T \hat{t}_1 \Big/ \left\|\hat{t}_1\right\|^2 \\ \beta_1 = F_0^T \hat{t}_1 \Big/ \left\|\hat{t}_1\right\|^2 \end{cases},$$

Step 3: Replace E_0 and F_0 with residual matrixes E_1 and F_1 and repeat the previous steps.

Generally, in PLS regression method, like the principal component analysis, the first l components $(l \le r)$ are necessary to acquire a regression model with preferable prediction, and the rest components are useless to establish the regression formulas. Cross validation is available to determine the number of principal components l.

In the process of iteration, a PLS regression model will be built each time with the residual $n-1$ observations, and the i th observation is put aside. Since the regression formulas are built by extracting h components, the predictive value of $y_j (j = 1, 2, ...p)$ is in the i th observation, $\hat{y}_{(i)j}(h)$ can be acquired by introducing the i th observation into the regression formulas. The above steps repeated by $i = 1, 2, ..., n$ will produce the predicted sum of squared error of the j th dependent variable $(j = 1, 2, ...p)$ under the circumstance of h components extracted, which is expressed as follows:

$$PRESS_j(h) = \sum_{i=1}^{n} (y_{ij} - \hat{y}_{(i)j}(h))^2 (j = 1, 2, ...p)$$

The predicted sum of squared error of $Y = (y_1, y_2, ..., y_p)^T$ is $PRESS(h) = \sum_{i=1}^{p} PRESS_j(h)$. The value of h to minimize $PRESS(h)$ is the needed component number.

CASE STUDY

PLS regression model is a new method of multivariate statistics analysis and its prominent characteristic is Multiple Linear Regression analysis, i.e., principal component analysis and canonical correlation analysis of variables are integrated and the regression modeling is realized in an algorithm so that the model can facilitate multivariate data analysis as well as simplify data structure and the correlation analysis of double variables. Besides, PLS regression model has its unique advantage when dealing with the problem of small sample size, independent variable and serious multicollinearity. However, as the PLS regression model can be expressed to be the regression equation of all the previous variables and the principal component it chooses will include all the variables, it cannot solve the more serious problem of correlations among variables, esp., it results in the unsatisfactory computational effect under the condition of independent variable and small sample size.

TABLE 1 : Physical training data

No.	Exercise Planning (x_1)	Self-efficacy (x_2)	Social Support (x_3)	Rotating (y_1)	Bending (y_2)	Bouncing (y_3)
1	91	36	50	5	162	60
2	89	37	52	2	110	60
3	93	38	58	12	101	101
4	62	35	62	12	105	37
5	89	35	46	13	155	58
6	82	36	56	4	101	42
7	67	34	60	6	125	40
8	76	31	74	15	200	40
9	54	33	56	17	251	250
10	69	34	50	17	120	38

In this section, the data of aerobics athletes' physical training is put into the PLS regression model, the ability of the three variables, i.e., exercise planning, self-efficacy and social support to explain and predict the exercise intention and behavior, along with the function in the process from intention to behavior are tested in the model. The respondents are 10 players in an aerobics team. The first group of variables is the quantitative X of exercise planning, self-efficacy and social support. There are two types of social supports in the study: social support from family and that from friends, each of which cannot predict exercise intention effectively. The second group of variables are training performance index Y including rotating, bending and bouncing. The raw data of physical training are shown in TABLE 1, correlation coefficient matrix is given in TABLE 2.

From the correlation coefficient matrix of 6 variables, it is obvious that a positive correlation exists between exercise planning and self-efficacy, a negative correlation exists between exercise planning, self-efficacy and social support, and a positive correlation exists between rotating, bending and bouncing.

In the process of PLS regression, t_1 and t_2 are extracted, with the cross-validation of $Q_2^{\,2} = -0.1969$. the regression formula of standardized variable $\overline{y_k}$ about component t_1 is as follows:

$$\overline{y_k} = r_{1k} t_1 + r_{2k} t_2, k = 1, 2, 3$$

TABLE 2 : Correlation coefficient matrix

No.	x_1	x_2	x_3	y_1	y_2	y_3
x_1	1	0.8702	-0.3658	-0.3897	-0.4931	-0.2263
x_2	0.8702	1	-0.3529	-0.5522	-0.6456	-0.1915
x_3	-0.3658	-0.3529	1	0.1506	0.225	0.0349
y_1	-0.3897	-0.5522	0.1506	1	0.6957	0.4958
y_2	-0.4931	-0.6456	0.225	0.6957	1	0.6692
y_3	-0.2263	-0.1915	0.0349	0.4958	0.6692	1

The PLS regression formula about t_1 is as follows:

$$\overline{y_k} = r_{1k}(w_{11}{}^{*}\overline{x_1} + w_{21}{}^{*}\overline{x_2} + w_{31}{}^{*}\overline{x_3}) + r_{2k}(w_{12}{}^{*}\overline{x_1} + w_{22}{}^{*}\overline{x_2} + w_{32}{}^{*}\overline{x_3})$$

$$= (r_{1k} w_{11}{}^{*} + r_{2k} w_{12}{}^{*})\overline{x_1} + (r_{1k} w_{21}{}^{*} + r_{2k} w_{22}{}^{*})\overline{x_2} + (r_{1k} w_{31}{}^{*} + r_{2k} w_{32}{}^{*})\overline{x_3}$$

The value of regression coefficient $r_h = (r_{h1}, r_{h2}, r_{h3})$ is in TABLE 3.

TABLE 3 : Regression Coefficient c(r_h)

k	**1**	**2**	**3**
r_1	0.3416	0.4161	0.1430
r_2	-0.3364	-0.2908	-0.0652

Therefore,

$$\overline{y_1} = -0.0778\overline{x_1} - 0.4989\overline{x_2} - 0.1322\overline{x_3} \quad ;$$

$$\overline{y_2} = -0.1385\overline{x_1} - 0.5244\overline{x_2} - 0.0854\overline{x_3} \quad ;$$

$$\overline{y_3} = -0.0604\overline{x_1} - 0.1559\overline{x_2} - 0.0073\overline{x_3}$$

Standardized variables $\overline{y_k}, \overline{x_k} (k = 1, 2, 3)$ are restored to $y_k, x_k (k = 1, 2, 3)$, the above formulas are changed into:

$$y_1 = 47.0197 - 0.0167x_1 - 0.8237x_2 - 0.0969x_3 \quad ;$$

$$y_2 = 612.5671 - 0.3509x_1 - 10.2477x_2 - 0.7412x_3 \quad ;$$

$$y_3 = 183.9849 - 0.1253x_1 - 2.4969x_2 - 0.0518x_3$$

In order to observe the effect of independent variables on deducing $y_k (k = 1, 2, 3)$, regression coefficients are shown on Figure 1. This figure demonstrates standardized data and regression formulas. It can be seen that self-efficacy plays an important role in interpreting 3 regression formulas. However, compared with rotating and bending, the regression formula of bouncing is not so ideal, and all the 3 variables respectively cannot account for the result very well.

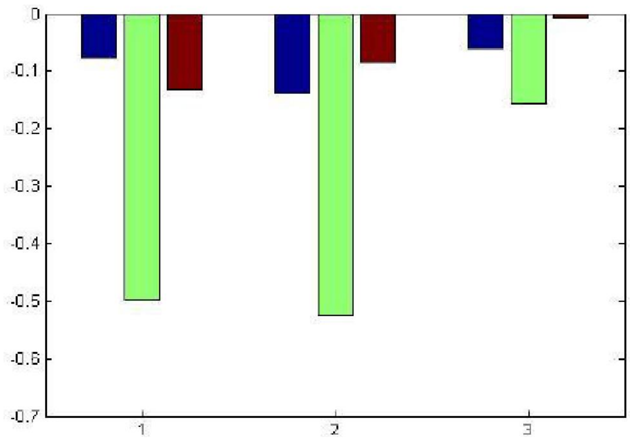

Figure 1 : A Histogram of regression coefficients

To better explain and predict the effect of self-efficacy on exercise behavior, 3 scatter diagrams about predictive and actual training results are drawn in Figure 2. The diagrams show that 3 correlations are all positive, with possible social expectation effect contained. For the sake of precision, all sample points are worked out in the diagrams, with the coordinates $(\hat{y}_{ik}, y_{ik})$, in which $\hat{y}_{ik}$ indicates the predictive value of the k th variable at the i th sample point. If the points are evenly distributed around the diagonal, the imitative effect is satisfactory.

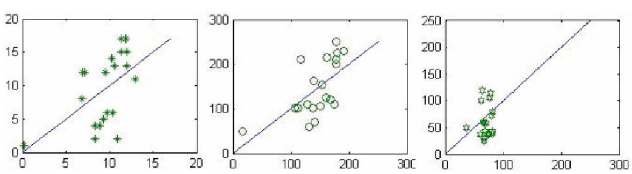

Figure 2 : Predictive and actual training results

The results of the model show that converting the exercise intention of aerobics athletes to their personal behaviors may be achieved by specific implementation intention (being aware of when, where and how to perform). Therefore, it is necessary to encourage aerobics athletes to have planned exercise behavior so as to strengthen and protect their intentions. In addition, individuals' experience of completing the task, i.e. the substitute experience, is an effective way to increase their confidence. Both the external social circumstances and the internal individual factors function greatly in the formation process of aerobics athletes' exercise behavior. The social support from family and friends, a variable paid attention to in the area of exercise behavior recently, is very important to help regulate individuals' physical activities.

CONCLUSION

The partial least squares regression model is the right combination of both the modeling method and the pattern recognition method which have very clear boundaries for a long time, and the model is a leap in the area of multivariate statistics analysis because regression modeling (Multiple Linear Regression), data structure simplification (Principal Component Analysis) and canonical correlation analysis can be realized in an algorithm. The results of this study support the prediction abilities of exercise motivation and behavior suggested by the variables of the planned behavior theory. The exercise behavior explained by planned behavior variables shows the significant prediction effect of intention. It is necessary to encourage aerobics athletes to have planned exercise behavior so as to promote and protect their intention and make it personal behavior. In addition, the experience of completing the task, i.e. the substitute experience, is an effective way to increase individuals' confidence. When studying the role of self-efficacy as mediate variable (whether the external social circumstances or the internal individual factor functions in the aerobics athletes' behavior), social support predictive behavior plays an important role. The study emphasizes self-efficacy as a mediator, not only does the training of self-efficacy have significant predictive effect towards exercise behavior intention, but also social support can reduce the contact pressure and improve the effort level and response capability.

REFERENCES

[1] A.Bandura; Perceived self-efficacy in cognitive-development and functioning. Educational Psychologist, **28**, 117-148 **(1993)**.

[2] L.Beiloeks, B.L.Bertenthal; Haste does not always make waste: Expertise, direction of attention, and speed versus accuracy in Performing sense or imotor skills. Psychonomic Bulletin & Review, **11**, 373-379 **(2004)**.

[3] L.Beiloeks, T.H.Carr; When high-powered people fail: Working memory and "Choking under Pressure" inmath. Psychology Science, **16**, 101-105 **(2005)**.

[4] Bing Zhang, Yan Feng; The Special Quality Evaluation of the Triple Jump and the Differential Equation Model of Long Jump Mechanics Based on Gray Correlation Analysis. International Journal of Applied Mathematics and Statistics, **40(10)**, 136-143 **(2013)**.

[5] Bing Zhang; Application of Mathematical Model of Evacuation for Large Stadium Building. Research Journal of Applied Sciences, Engineering and Technology, **5(04)**, 1432-1440 **(2013)**.

[6] M.W.Eysenck, N.Derakshan, R.Santos; Anxiety and cognitive Performance: Attentional control theory. Emotion, **7**, 336-353 **(2007)**.

[7] K.Hanulton, K.M.White; Extending the theory of Planned behavior: The role of self and social influences in predicting adolescent irregular mode rate to vigorous physical activity. Journal of Sport& Exercise Psychology, **30**, 56-74 **(2008)**.