INFORMATION
TECHNOLOGY JOURNAL

Research on Application of Matlab Uniform Design in Psychological Experiment of Factors Affecting Reading Comprehension

Wenjun Cong

School of Education, Huaibei Normal University, 235000, Huaibei, China

Abstract: Experimental psychology is an essential primary discipline for psychological teaching and research, with the continuous development of science and technology, measurement techniques and computer data simulating techniques have a great power to implement the experimental design and the design method of experiment is also very important. In this study, the uniform design experiment is adopted to study the factors affecting reading comprehension in psychology, this study first gives four factors affecting reading comprehension, the level of the four factors is technically divided into five grades, the uniform design of psychological experiments is then analyzed, finally carry out regression analysis on the factors affecting reading comprehension by using the uniform design's experimental data of five testers, thus the effect of factors affecting reading comprehension is obtained, which provides a theoretical basis for the experimental design and data analysis of psychology, expecting a more distant development of experimental psychology.

Key words: Uniform design, Regression model, matrix, least square method, matlab programming

INTRODUCTION

The complexity and uncertainty of psychological research itself is far beyond physics and other scientific researches, which undoubtedly makes the empirical observation of experimental psychology become very difficult, the research objects of psychology are human who possess their own subjective will and complex feeling, to make experimental psychology meet the standards of scientific method in terms of empirical observation, we must have a set of effective and stable methods of observation and processing observation value (Zhang, 2013).

Many people have made effects to study the experimental psychology and uniform design experiment, it is these people's efforts that the two can be combined together, making the design methods of psychological experiment have been greatly improved, in recent years, some scholars put forward their own views, among which are: Fan (2009) proposed the uniform design method that the number of experiments is equal to those of processing level, this method is suitable for exploratory psychological experiments, particularly for detecting the interaction of a number of factors, this design method can also process continuous variables, discrete variables, it can be applied in psychological experiments; according to the multivariate nonlinear least square principle, Wang (2010) used the MATLAB programming method to fit a set of empirical formula of uniform design's experimental results, conclusions and theoretical analysis drawn from fitting

formula are consistent with other studies' experimental results; Zeng Z.j. mentioned that in 1978, Ministry of Space Industry had raised a five- factor test, the number of level of each factor is greater than 10 and the number of experiments is required to no more than 50 times, if the orthogonal design can not accomplish the experiment, there is a need to adopt uniform design (Cui, 2013).

On the basis of previous studies, the uniform design experiment is adopted in this study to analyze the factors affecting the difficulty of reading comprehension, the nonlinear regression model and Matlab software are also used, expecting to provide some suggestions for experimental psychology through the research content and results in this study.

DESIGN OF PSYCHOLOGICAL EXPERIMENTS

Factors affecting the difficulty of reading comprehension: In this study, the factors affecting the difficulty of reading comprehension are summered as vocabulary density, sentence length, theme familiarity, beauty degree of rhetoric, in order to digitize the four factors, they can be divided into different levels according to the following method:

- Vocabulary density is defined as the ratio of vocabulary number to the total number, its level is divided into five grades and it is divided according to the membership function shown in Eq. 1:

$$A_{level} = \begin{cases} 1 & k \le 1/25 \\ 2 & 1/25 < k \le 1/20 \\ 3 & 1/20 < k \le 1/15 \\ 4 & 1/15 < k \le 1/10 \\ 5 & 1/10 < k \le 1/5 \end{cases} \qquad (1)$$

In Eq. 1, A_{level} indicates the level of vocabulary density, the difficulty increases from 1-5, k indicates the ratio of vocabulary number to the total number

- Sentence length refers to the average number of words of sentences in an article. Sentence length is divided into five grades; it is divided according to the membership function shown in Eq. 2:

$$B_{level} = \begin{cases} 1 & n \le 5 \\ 2 & 5 < n \le 15 \\ 3 & 15 < n \le 20 \\ 4 & 20 < n \le 25 \\ 5 & 25 < n \le 30 \end{cases} \qquad (2)$$

In Eq. 2, B_{level} indicates the level of sentence length, the difficulty increases from 1-5, n indicates the average number of words of sentences in an article

- Theme familiarity is the association degree of article theme and the tester's life, since the technical constraints involved to theme familiarity are difficult to quantified, the expert scoring method is adopted to determine the extent of the theme, divide the level of factors reflecting the difficulty of reading comprehension into five grades, which is denoted by C_{level}, the difficulty levels from small to large are 1, 2, 3, 4, 5

- The beauty degree of rhetoric is assessed by experts, according to common sense, the more beautiful the rhetoric is, the easier the readers can understand, divide the beauty degree of rhetoric into five grades, the higher the grade is, the higher the beauty degree is, otherwise, the lower the difficulty level is, therefore, the difficulty levels from small to large are 1, 2, 3, 4, 5, which is denoted by D_{level}.

Design of psychological experiments: In order to ensure the experimental design is feasible and easy to carry out, the uniform design is adopted in this study, this design requires for ten people, each one reads five articles, select the appropriate uniform design table according to the number of factors and levels, the number of level of each factor is 5, so the five-level uniform design table is selected in this study, which is U_5 (5^4), combined with U_5 (5^4) and the selected table, the experimental arrangement can be shown in Table 1.

Experimental arrangement in Table 1 can only deal with four variables and one interaction, therefore, it

Table 1. U_5 (5^4) Design of experiment $5 \times 5 \times 5 \times 5$

Experimental processing	Affecting factors and levels			
	A	B	C	D
1	A1	B1	C1	D1
2	A2	B2	C2	D2
3	A3	B3	C3	D3
4	A4	B4	C4	D4
5	A5	B5	C5	D5

Table 2: U_{10} (5^5) Design of experiment $5 \times 5 \times 5 \times 5$

Experimental processing	Affecting factors and levels			
	A	B	C	D
L1	A1	B2	C2	D3
L2	A1	B3	C4	D5
L3	A2	B5	C1	D2
L4	A2	B1	C3	D5
L5	A3	B2	C5	D2
L6	A3	B4	C1	D4
L7	A4	B5	C3	D1
L8	A4	B1	C5	D4
L9	A5	B3	C2	D1
L10	A5	B4	C4	D3

requires for a uniform design table that can process more experiments, following the U_5 (5^4) principle, select U^*_{10} (10^8), if following the U^*_{10} (10^8) table, similarly select four factors of A, B, C, D, compose the levels of these four factors to constitute 10 levels, make {1,2}=>1£¬{3,4}=>2£¬{5,6}=>3£¬{7,8}=>4£¬{9,10}=>5 in each list, thus to obtain U_{10} (5^5), which is shown in Table 2.

PROCESSING MODEL OF UNIFORM DESIGN'S EXPERIMENTAL DATA

Fitting theory and matrix expression of multivariate nonlinear model: The least square method is adopted in this study to fit the multivariate linear model, if there is a set of data whose independent variables are x, y, z,... and dependent variable is w and the known function f_j (j = 1, 2, ..., m) is selected to construct the regression equation, there is a need to solve the coefficient of regression equation k_j (j = 1, 2, ..., m), making the residual sum of squares of regression value and experimental value □□of each set of data minimum, which is shown in Eq. 3:

$$Q(k_1, k_2, \cdots, k_m) = \sum_{i=1}^{n} \left[\hat{w}_i(k_1, k_2, \cdots, k_m) - w_i \right]^2 = \min \qquad (3)$$

In Eq. 3, Q indicates the residual sum of squares, $\hat{w}_i$ indicates the ith regression value, w_i indicates the ith experimental value, it is concluded from the necessary conditions when multivariate function obtains extremes that when there is a minimum value for Q, the Eq. 4 must be existed:

Table 3: Simulation data of uniform design U_{10} (5^5)

Testers	Experimental processing									
	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10
P1	21	8	19	13	12	10	11	13	15	7
P2	22	9	17	14	11	9	9	12	16	6
P3	24	7	15	11	10	8	10	10	17	5
P4	25	6	18	9	13	6	12	9	14	8
P5	20	9	16	10	9	9	8	11	13	4
Average	22.4	7.8	17	11.4	11	8.4	10	11	15	6

Table 4: List of estimation of rregression coefficient after the revise of the number of independent variables

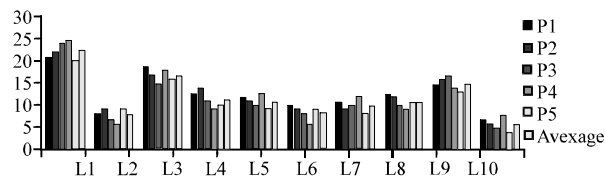| Parameter	Estimate	Asymptotic Std. error	Asymptotic 95% Confidence interval	
			Lower	Upper
M	3.372	4.678	-9.615	16.359
k_1	28.984	45.830	-98.260	156.227
K_4	1.253	1.517	-2.958	5.464
K_5	-2.812	1.395	-6.686	1.062
K_6	14.954	9.642	-11.818	41.725
K_7	11.275	17.988	-38.667	61.217

Fig. 1: Distribution diagram of five tester's experimental data

$$\frac{\partial Q}{\partial k_i} = 0 (i = 1, 2, \cdots, m) \tag{4}$$

So, the linear equations of k_1, k_2,..., k_m shown in Eq. 5 can be obtained:

$$\begin{cases} \sum_{i=1}^{n} f_1(x_i, y_i, z_i \cdots) \left[\sum_{j=1}^{m} k_j f_j(x_i, y_i, z_i \cdots) - w_i \right] = 0 \\ \cdots \\ \cdots \\ \sum_{i=1}^{n} f_m(x_i, y_i, z_i \cdots) \left[\sum_{j=1}^{m} k_j f_j(x_i, y_i, z_i \cdots) - w_i \right] = 0 \end{cases} \tag{5}$$

Assuming R is shown in Eq. 6, K is shown in Eq. 7, W is shown in Eq. 8, the Eq. 5 can be expressed in the form of the matrix shown in Eq. 9:

$$R = \begin{pmatrix} f_1(x_1, y_1, z_1 \cdots) & f_2(x_1, y_1, z_1 \cdots) & \cdots & f_m(x_1, y_1, z_1 \cdots) \\ f_1(x_2, y_2, z_2 \cdots) & f_2(x_2, y_2, z_2 \cdots) & \cdots & f_m(x_2, y_2, z_2 \cdots) \\ \vdots & \vdots & \vdots & \vdots \\ f_1(x_n, y_n, z_n \cdots) & f_2(x_n, y_n, z_n \cdots) & \cdots & f_m(x_n, y_n, z_n \cdots) \end{pmatrix} \tag{6}$$

$$K = (k_1, k_2, \cdots, k_m)^T \tag{7}$$

$$W = (w_1, w_2, \cdots, w_n)^T \tag{8}$$

$$R^T R K = R^T W \tag{9}$$

If f_1 (x, y, z ...), f_2 (x, y, z...), ..., f_m (x, y, z...) is linear unrelated, $R^T R$ is reversible, the Eq. 9 of K have a unique solution, in this study, the Matlab programming method is used to solve the Eq. 9, thus to achieve the nonlinear least square fitting of the multivariate data, the regression coefficient is calculated shown in Eq. 10:

$$K = (R^T R)^{-1} R^T W \tag{10}$$

Analysis of uniform design's experimental data: Experiment designed in Table 2 requires the use of regression analysis method, then choose the appropriate dependent variable to establish the nonlinear regression model through the analysis results, the experimental data of five testers is shown in Table 3.

Distribution of the data in Table 3 is shown in Fig. 1.

Model selection and analysis: It can be learned from the data distribution in Fig. 1 that variables affecting the reading comprehension can be summarized as 8, which are A, B, C, D, A×B, A×C, A×D, C×D, assuming the eight variables as independent variables, W indicates the difficulty of reading comprehension and k_1, k_2,..., k_8 indicate the coefficients of these 8 independent variables, thus the multivariate nonlinear regression equation is shown in Eq. 11:

$$W = m + k_1 A + k_2 B + k_3 C + k_4 D + k_5 A \times B \\ + k_6 A \times C + k_7 A \times D + k_8 C \times D \tag{11}$$

In Eq. 11, m indicates a constant term.

Using matlab software for data analysis can obtain the regression equation shown in Eq. 12:

$$W = 11.849 + 44.937A - 0.187B + 0.591C - 3.980D \\ - 1.789A \times B - 29.820A \times C + 53.952A \times D + 0.898C \times D \tag{12}$$

The test parameters of the regression equation in Eq. 12 are shown in Eq. 13:

$$\begin{cases} F(9,1) = 3831.188 \\ p < 0.001 \\ R^2 = 0.999 \end{cases} \quad (13)$$

It can be learned from the coefficient's characteristics of regression equation model in Eq. 12 that, the effect of sentence length, theme familiarity, crossing factors of theme familiarity and beauty degree of rhetoric on reading comprehension is very low, so these three variables can be removed, select five variables of A, D, A×B, A×C, A×D as independent variables to re-fit the nonlinear regression model, similarly use Matlab software to program, thus to estimate the regression coefficient, which is shown in Table 4.

Therefore, the revised regression equation is shown in Eq.14:

$$W = 3.372 + 28.984A + 1.253D \\ +14.954A \times C + 11.275A \times D \quad (14)$$

The goodness of fit of Eq. 14 R^2 is shown in Eq. 15:

$$R^2 = \frac{1 - \text{Residual SS}}{\text{Corrected SS}} = 0.885 \quad (15)$$

It can be learned from Eq. 15 that in case of a higher goodness of fit, utilizing the above five factors can reflects the difficulty of reading comprehension.

Research results: It can be learned from the above analysis of simulated data model that, the vocabulary density has the greatest impact on reading comprehension, followed by beauty degree of rhetoric and vocabulary density has a significant degree of interaction on sentence length, theme familiarity and beauty degree of rhetoric.

CONCLUSION

The factors affecting the difficulty of reading comprehension is analyzed in this study, summered as

four factors and each related factor has five grades; the design method of psychological experiment is also studied in this study, the uniform design experiment is carried out according to the research object-factors affecting reading comprehension in this study; the experimental data's research model-multivariate nonlinear regression model is also analyzed in this study, as well as the solving principle of the model; the experimental data of five testers are analyzed and then on the basis of the set of data, use Matlab software for programming, thus to obtain the nonlinear regression equation, through analyzing the initial regression equation, the regression equation of five variables is then concluded; the effect of factors affecting reading comprehension ranked from small to large is vocabulary density>beauty degree of rhetoric>sentence length>theme familiarity, by analyzing, it can be concluded that vocabulary density has a significant interaction with the remaining three factors.

REFERENCES

Cui, C., 2013. Application of mathematical model for simulation of 100-meter race. Int. J. Applied Math. Stat., 42: 309-316.

Fan, H.Y., 2009. Discussion on application of uniform design in psychological experiments. Adv. Psychol. Sci., 17: 233-239.

Wang, H.P., 2010. Matlab-based multivariate nonlinear least square fitting of uniform design's experimental data. Chem. Eng. Equipment, 9: 29-32.

Zhang, B., 2013. Dynamics mathematical model and prediction of long jump athletes in Olympics. Int. J. Applied Math. Stat., 44: 422-430.