ATLANTIS PRESS
Advances in Social Science, Education and Humanities Research, volume 333
Humanities and Social Sciences: Novations, Problems, Prospects (HSSNPP 2019)

Mathematical Modeling in Multilevel Educational Programs

Nonlinear Models for Test Validity and Reliability

Ivanov D.V.
Department of Higher Mathematics and Economic and Mathematical Methods
Samara State University of Economics
Samara, Russia
dvi85@list.ru

Merkulova L.P.
Department of foreign languages and Russian as a foreign language
Samara National Research University named after academician S.P. Korolev
Samara, Russia
aimerk@yandex.ru

Nelyubina E.G.
Department of Chemistry, Geography and Methods of Teaching
Samara State University of Social Sciences and Education
Samara, Russia
nelubina.elena@pgsga.ru

Panfilova L.V.
Department of Chemistry, Geography and Methods of Teaching
Samara State University of Social Sciences and Education
Samara, Russia
panfilova@pgsga.ru

Bobkova E.Yu.
Moscow State University of Technologies and Management
Moscow, Russia
bobkova_e_yu@e-sceo.ru

***Abstract*—The quality of education depends to a large extent on the quality of educational programs. Educational institutions develop educational programs to improve the quality of education. This involves not only monitoring changes that occur in society but also taking forecast-based steps to make educational programs predictive. The major tool for improving educational programs is student assessment. It, in turn, involves testing—the only truly practicable tool to evaluate knowledge, helping teachers assess and manage the effectiveness of instruction. For a test to assess a student's knowledge adequately, it should have the necessary characteristics. The most important of those are reliability and validity. Variables in educational tests, especially multilevel ones, often show nonlinear associations. Because of this, the use of correlation measures of association may lead to inaccurate—and sometimes incorrect—results. This paper proposes using correlation ratios to assess the validity and reliability of tests.**

Keywords— innovation modeling, hereditary models, fractional calculus, errors in variables, least-squares method

I. Introduction

The quality of education—its ability to satisfy the evolving needs of society, the government, and the individual—depends to a large extent on the quality of educational programs. The government places a constant emphasis on developing new approaches to creating educational standards to improve the educational programs in place at seats of learning. At the same time, with the advent of new-generation educational standards, there are now more opportunities for teachers to contribute to developing educational programs.

It is important to understand, as part of modeling those programs, that because of the inertia of education systems, improving the quality of education should include making it predictive. This is especially relevant for vocational education. Vocational courses last long enough to result, should drastic social and economic changes occur, in an imbalance between the qualifications of graduates and the qualification requirements from the industry in question.

Thus, improving the quality of education by developing the educational programs at seats of learning involves not only monitoring societal changes but also taking forecast-based project steps to make educational programs (their values, structures, etc.) predictive. The forecast-based development of educational programs relates to potential changes in the subject matter, objectives, types, and methods of education and student assessment as well as in the interaction between students and teachers and in the organizational and academic influence on that interaction. How those changes are interrelated is important to consider.

II. Educational Measurement

Testing as an assessment tool is becoming an essential part of instruction. Although it has some drawbacks, testing is the only truly practicable tool to evaluate knowledge, helping teachers assess and manage the effectiveness of instruction [1, 2].

The quality of testing materials is crucial for student assessment. To avoid inadequate assessment, one should use only statistically sound testing materials that are reliable and valid enough.

For a test to assess a student's knowledge adequately, it should have the necessary characteristics. The most important of those are reliability and validity.

Reliability is a criterion that determines the quality of a test, ensuring measurements are accurate and test results are immune to random extraneous factors.

The more consistent the test results for the same person when they retake the test or its alternative (a parallel test), the more reliable the test.

Validity is the ability of a test to yield the results that are consistent with the objective pursued. Test validity shows how well the test does what it is designed for.

The following methods are used to evaluate test reliability:

—The retest method

—The parallel-form method

—The split-halves method

Reliability evaluation involves computing the correlation between two sets of results for the same test or for two parallel forms. The higher the correlation, the more reliable the test is. A test has a good reliability coefficient when that indicator falls within the range $0.8 < r < 1$.

The retest method consists in administering a test twice to the same group. The method involves computing the correlation of test-takers' individual scores from the first and the second testing. The retest method is not very convenient because of the marked effect of the time factor: the interval between the tests should be neither too long (the student's educational achievements progress to a different level over time), nor too short (the student may remember the test items and the answers).

The parallel-form method involves testing the same group twice with tests that have identical contents and structures and that include test items with the same levels of difficulty and discrimination power. The major difficulty inherent in using that method is having to devise a second test identical to the first one and also to prove that they are identical, a procedure that is too complicated.

The split-halves method is convenient to use in practice since it does not require a repeat testing. The method splits a test in two subtests and assumes that the subtests are parallel. The test items are divided into odd-numbered items on one subtest and the even-numbered on the other.

Validity is verified by comparing the student's test results with the results of the expert assessment of the student's knowledge with a different method—an oral questionnaire, a traditional test or examination—or by comparing those results with the student's current academic performance.

Validity and reliability are usually computed with the Pearson correlation coefficient:

$$r_{xy} = \frac{\operatorname{cov}(x, y)}{\sigma_x \cdot \sigma_y}, \qquad (1)$$

where $\operatorname{cov}(x,y) = \overline{y \cdot x} - \bar{y} \cdot \bar{x}$, $\sigma_x^2 = \overline{x^2} - \bar{x}^2$,

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i,\ \bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i,\ \overline{xy} = \frac{1}{N}\sum_{i=1}^{N} x_i y_i,\ \overline{x^2} = \frac{1}{N}\sum_{i=1}^{N} x_i^2 .$$

If the variables x and y are not distributed normally, Spearman's and Kendall's rank correlation criteria can be used.

Many coefficients based on correlation variables find application today. Notice that various correlation coefficients identify only the linear association between variables. With a nonlinear association, a correlation coefficient yields lower estimates of association or shows no association at all.

Fig. 1 and 2 show scatter diagrams and the Pearson correlation coefficient.

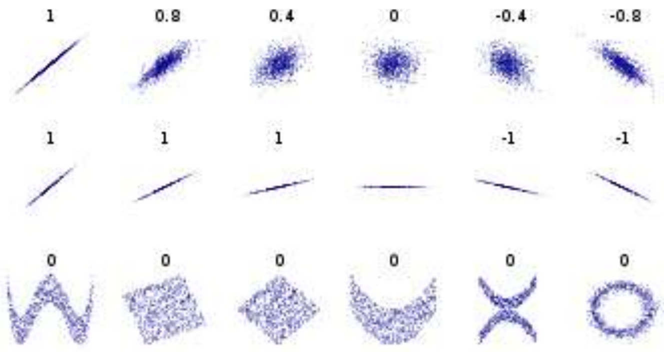

Fig. 1. Scatter diagrams and Pearson correlation coefficient. [3]

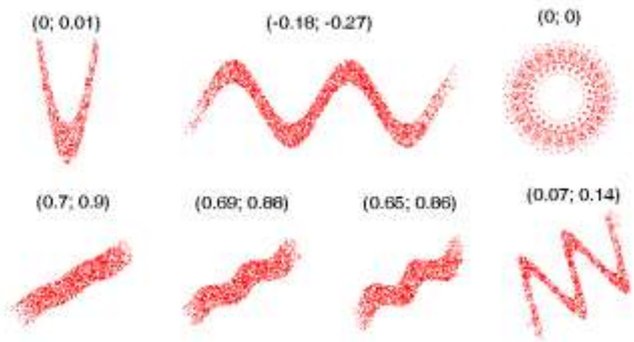

Fig. 2. Scatter diagrams and Spearman and Kendall correlation coefficient. [4].

The Spearman and Kendall correlation coefficient is more resistant to outliers; but when the nonlinearity is highly pronounced, the coefficient fails as well. Spearman's correlation coefficient value is usually greater than Kendall's correlation coefficient.

III. Methods for Estimating Nonlinear Functional Relations

Reference [5, 6] reports that in various psychometric studies, of which educational tests are a special case, variables often show a nonlinear association. Because of this, the use of correlation measures of association may lead to inaccurate—and sometimes incorrect—results.

It is noteworthy that test theory often concerns itself with looking into the problem of determining associations for multidimensional variables.

For instance, in the problem of assessing the validity of test items, there should be a positive functional relation between the number of correct answers to a question and the score—preferably with no relation between the test items. Multilevel tests do not meet the condition of linearity between the score and the number of correct answers to a question because the test items have different levels of difficulty that require different sets of skills.

The problem of determining nonlinear associations between variables is often solved through regression analysis.

A. *The all-possible-regressions method*

For the identification problem, functions to renew $\mathrm{E}[y/x]$ can, generally speaking, be selected arbitrarily if there is no prior information about $\mathrm{E}[y/x]$. But once the choice is made, the optimal type of relationship for the class of functions can be determined with one of the following methods:

- The exclusion method
- The inclusion method
- The stepwise regression method
- The stagewise regression method

These methods are modifications of the all-possible-regressions method, which requires consecutively fitting all possible sets of variables and possible dependencies on the variables considered. The complexity of these methods for nonlinear multivariate functions is evident.

B. *Correlation ratio*

With a general nonlinear association, a correlation ratio can be used as a measure of association [7-9]. The correlation ratio is used to process data received from complicated technical facilities and medical data.

Let us express σ_Y^2 as

$$\begin{aligned} \sigma_Y^2 &= \mathrm{E}[(Y - M[Y])^2] = \\ &\mathrm{E}[(Y - M[Y/X])^2] + \mathrm{E}[(M[Y/X] - \mathrm{E}[Y])^2] = \\ &= \mathrm{E}[(Y - \bar{y}(x))^2] + \mathrm{E}[(\bar{y}(x) - \mathrm{E}[Y])^2] = \sigma_{Y/X}^2 + \delta_{Y/X}^2 \end{aligned} \quad (2)$$

Here, the first summand

$$\sigma_{Y/X}^2 = M[(Y - \bar{y}(x))^2] = M[(Y - M[Y/X])^2]$$

is the dispersion of the attribute Y in relation to the function of regression of Y on X, which is called residual dispersion or conditional dispersion. It is a measure of scatter of the attribute Y around the regression line. Residual dispersion is that portion of the resultant attribute's scatter, which cannot be explained by the impact of the observable attribute. Residual dispersion can serve to assess how accurately the type of regression function (the model regression equation) is selected.

If X and Y are independent, then all conditional mathematical expectations coincide with unconditional ones and $\sigma_{Y/X}^2 = \sigma_Y^2$.

The second summand $\delta_{Y/X}^2$ is the dispersion of the function of regression of Y on X in relation to the mathematical expectation of the attribute Y, and it measures the effect of the attribute X on Y. It can be used to assess the strength of the association between X and Y. Let us introduce the correlation ratio

$$\eta_{Y/X}^2 = \frac{M[(\bar{y}(x) - M[Y])^2]}{\sigma_Y^2} = \frac{\delta_{Y/X}^2}{\sigma_Y^2} \quad (3)$$

Dividing both parts of (2) by σ_Y^2, we obtain a correlation ratio of the form

$$\eta_{Y/X}^2 = 1 - \frac{\sigma_{Y/X}^2}{\sigma_Y^2}. \quad (4)$$

From (4) it follows that

1) $0 \le \eta_{Y/X}^2 \le 1$, because $\sigma_{Y/X}^2$ is a component of σ_Y^2 and so $\sigma_{Y/X}^2 \le \sigma_Y^2$;

2) $\eta_{Y/X}^2 = 1$ only if $\sigma_{Y/X}^2 = 0$; that is, when all the other factors have no effect and the entire spread concentrates on the regression curve $\bar{y}(x)$;

3) $\eta_{Y/X}^2 = 0$ only if $\sigma_{Y/X}^2 = \sigma_Y^2$; that is, $\bar{y}(x) = \mathrm{E}[Y] = const$. Here, the regression line Y on X is a horizontal straight line that traverses the distribution center. Then the variable Y is not correlated with X; that is, the correlation coefficient equals zero, $r = 0$.

If the regression of Y on X is linear—that is, the regression line is straight,

$$y(x) = M[Y] + r\frac{\sigma_Y}{\sigma_X}(x - M[X])$$

—then $\sigma_{Y/X}^2 = \sigma_Y^2(1 - r^2)$ and $\eta_{Y/X}^2 = r^2$.

If $|r| = 1$, then Y and X are linked by a precise linear relation; but if $\eta_{Y/X}^2 = r^2 < 1$, then there is no functional relation between Y and X. The precise functional relation between Y and X other than a linear one takes place if and only if $r^2 < \eta_{Y/X}^2 = 1$.

The values of η^2 within the range of $0 < \eta^2 < 1$ indicate how tightly points cluster around the regression curve regardless of what type it is (regardless of the form of association). The correlation ratio η^2 is linked to the correlation ratio r^2 by the ratios $0 \le r^2 \le \eta^2 \le 1$. If there is a linear relation between the variables, $r^2 = \eta^2$. The difference $\eta^2 - r^2$ can be used as an indicator of the nonlinear association between the variables—that is, the deviation of the regression from linearity.

Note that when there are several function values for a single argument, the correlation ratio shows that there is no association. That type requires other measures of association and is beyond the scope of this paper.

IV. Simulation Results

Input data used to determine test quality are the test results for a sample of students, and those results are presented in a disordered results matrix whose columns correspond to the test items; and lines, to the students' last names. The elements of the disordered test matrix are the student's results for completed test items, assessed with a continuous system in the range [0,1].

Based on the disordered matrix, an ordered matrix is made whose data are the input for further computations. The disordered results matrix is used to identify invalid test items (items that none of the students completed and items that all the students completed). Those test items are not included in the ordered matrix.

The remaining test items are put in order as follows: the lines of the matrix are arranged by the score for all items completed by each student in ascending order from top to bottom, and the columns are arranged by the score for each item completed by all the students in descending order from left to right.

Table I lists the coefficients of correlation between test items and scores. Table II lists the correlation ratios for test items and scores. The method for calculating correlation ratios can be found in [10].

The results in Table I show that item 5 is correlated loosely with the score while test items 2 and 4 are correlated strongly with each other.

The results in Table II show that test item 5 is functionally dependent on the score while the items 2 and 4 and 5 and 6 are highly interdependent.

V. Conclusion

This paper showed that correlation ratios can be successfully used in assessing the quality of tests in multilevel educational programs. Correlation ratio–based indicators of test quality complement and do not conflict with the traditional quality indicators.

We intend to advance this research by exploring whether it is possible to use in educational assessment the partial and multiple correlation ratios equivalent to partial and multiple correlation coefficients.

TABLE I. Matrix of Pearson Correlation Coefficients

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**S**
1	1								
2	0.23	1							
3	0.32	0.25	1						
4	0.15	**0.92**	0.22	1					
5	0.33	0.22	0.11	0.37	1				
6	0.14	0.19	0.25	0.36	0.39	1			
7	0.17	0.19	0.41	0.12	0.33	0.21	1		
8	0.37	0.24	0.21	0.23	0.34	0.27	0.29	1	
S	0.82	0.60	0.63	0.87	**0.27**	0.78	0.73	0.65	1

TABLE II. Matrix of Correlation Ratios

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**S**
1	1								
2	0.23	1							
3	0.32	0.25	1						
4	0.15	**0.92**	0.22	1					
5	0.33	0.22	0.11	0.37	1				
6	0.14	0.19	0.25	0.36	0.39	1			
7	0.17	0.19	0.41	0.12	0.33	0.21	1		
8	0.37	0.24	0.21	0.23	0.34	0.27	0.29	1	
S	0.82	0.60	0.63	0.87	**0.27**	0.78	0.73	0.65	1

References

[1] D.D. Christopher and O. Bulut, "Handbook of Educational Measurement and Psychometrics Using," R. Chapman and Hall/CRC: Boca Raton, 2018.

[2] M.I. van Blerkom, "Measurement and statistics for teachers," Taylor and Francis: New York, 2009.

[3] K. Pearson, "Mathematical contributions to the theory of evolution. III. Regression, heredity, and panmixia," Phil. Trans. Roy. Soc. London A, vol. 187, pp. 253-318, 1896.

[4] M.G. Kendall, "A new measure of rank correlation," Biometrika, vol. 30, pp. 81–93, 1938.

[5] E.M. Tucker-Drob, K.P. Harden and E. Turkheimer, "Combining nonlinear biometric and psychometric models of cognitive abilities," Behavior genetics, vol. 39(5), pp. 461–471, 2009.

[6] D.J. Bauer, "The Role of Nonlinear Factor-to-Indicator Relationships in Tests of Measurement Equivalence," Psychological Methods, vol. 10(3), pp. 305-316, 2009.

[7] L.P. Ayres, "The Correlation Ratio," Journal of Educational Research, vol. 2, pp. 452–456, 1920.

[8] J. P. Shaffer and M.V. Gillo, “A Multivariate Extension of the Correlation Ratio,” Educational and Psychological Measurement, vol. 34(3), pp. 521–524, 1974.

[9] A.R. Sampson, “A multivariate correlation ratio,” Statistics & Probability Letters, vol. 2(2), pp. 77-81, 1984.

[10] A. R. Crathorne, “Calculation of the Correlation Ratio,” Journal of the American Statistical Association, vol. 18, No. 139, pp. 394-396, September 1922.