Adult and Higher Education (2023)
Clausius Scientific Press, Canada

DOI: 10.23977/aduhe.2023.050420
ISSN 2523-5826 Vol. 5 Num. 4

Mathematical Statistics and Big Data Analysis on Intelligent Teaching Evaluation Model

Xiaobo Wen

Sichuan Minzu College, Kangding, China
140339@scun.edu.cn

Keywords: Trimmed mean; teaching evaluation; big data analysis; reliability and validity analysis; factor analysis

Abstract: The advent of the era of big data has had a profound impact on teaching evaluation. With the help of statistical knowledge, this paper starts with data preprocessing such as trimmed mean and score standardization, and further discusses the application of big data analysis, data consistency and correlation, reliability and validity in teaching evaluation. Based on mathematical statistics and big data analysis, a systematic intelligent teaching evaluation model is constructed.

1. Introduction

The quality of teaching is not only determined by teachers unilaterally, but is the result of a combination of factors. The main factors affecting the quality of teaching are teachers' teaching quality, knowledge reserve, teaching operation management and students' quality. Teaching quality evaluation is a common form of obtaining teaching feedback. Teaching evaluation can improve teachers' teaching quality and ensure high-quality training of students. As far as schools are concerned, in the past, teachers in colleges and universities were promoted, raised, and rewarded based on their scientific research achievements. Teaching only needed to complete the corresponding class hours. As a result, college teachers paid more attention to scientific research than teaching. Build a teaching evaluation system, establish quantitative teaching effects and attach importance to teaching to improve teaching quality. Teaching evaluation can be used as a reference for the quality of teachers' teaching, allowing teachers to self-reflect on teaching, optimizing educational methods and methods, so that college teachers can return to their origins, and teachers teach well and students learn knowledge well.

Statistics is good at collecting, sorting, analyzing and inferring data, and can be widely used in teaching evaluation. This paper applies the relevant knowledge of statistics to teaching evaluation, explores some feasible teaching evaluation data processing and analysis methods, and analyzes some problems with practical cases, and draws some feasible theories and analysis methods. Relying on these mathematical statistics and big data analysis methods, an intelligent teaching evaluation model system is constructed[1-2].

2. Data preprocessing

The data obtained in the teaching evaluation often have extreme values, which cannot be directly used for data analysis. It is necessary to use data preprocessing technology to do some preliminary operations on the data. In this paper, the trimming mean processing method is introduced in the teaching evaluation to preprocess the data. The data obtained in the teaching evaluation sometimes needs to be used for horizontal comparison, the combination of relative and absolute data is required, and the standardized method of scores needs to be used to preprocess the data.

2.1 The use of trimmed mean

Students' teaching evaluation of teachers is a large part of teacher's teaching evaluation, but students' scores are not stable, but it is not good for understanding the teaching situation of teachers. For example, after receiving 100 students' teaching evaluations for this course in a teaching evaluation, the average score, is often used to estimate the teacher's teaching. However, the average value is easily affected by extreme values, and sometimes does not truly reflect the average value of the data. We can use the trimmed mean to measure the average level of teachers, that is, remove some of the highest scores and some of the lowest scores, and use the remaining data to calculate the trimmed mean.

In statistics, the tail cutting rate is generally between 5% and 10%. If there are too many tail cuttings, too many samples will be lost, and the excessive loss of data will distort the average result. Moreover, in order to maintain a balance, the method of equal tail trimming is generally adopted, that is, the data of the upper and lower trimming tails are the same, so that whether the data is symmetrical or not, the mean value of the tail trimming can truly reflect the average level of the data.

Especially when there are extreme values in the data, the influence of extreme values is removed, which makes the tail-cut mean more robust as an estimate of students' scores. So how to determine the tail cutting rate, there are two ways. The simpler way is to directly fix the tail cutting rate, for example, 5% of the upper and lower parts are removed, and 10% of the whole is removed. For example, 100 students are selected above. Under the condition of the tail cutting rate, 5 of the highest score and the lowest score are removed. Although the preprocessing of data in this way is relatively simple, the control of the resection rate of 5% of the upper and lower resections is also highly subjective and sometimes lacks factual basis. Another way is to use the characteristics of the data itself to control the tail cutting rate, that is, the 3s principle to process the data.

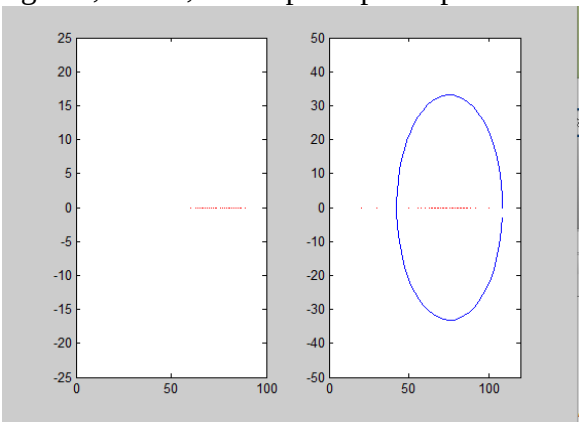

Figure 1 Ordinary data and 3s standard truncation

The original data generally obeys the normal distribution or approximately obeys the normal distribution. The sample mean and sample variance can be calculated from the original data, with the mean as the center and the data outside the radius of 3s on the left and right are invalid data and can

be removed. Then use the remaining data to calculate the trimmed mean, as shown in Figure 1. The range of scores in the teaching evaluation is not large, and it is reasonable to use the characteristics of this data to determine the mean of the trimming rate. However, this kind of tail trimming using the characteristics of the data itself is often not necessarily equal tail trimming, that is, the number of data removed from the top and bottom is not the same, which is also the limitation of this method. Therefore, it is necessary to think dialectically when doing the mean value of the trimming of the teaching evaluation data, which method should be used for calculation. The following is an example of the educational test scores of 100 students collected, and the tail-cut mean is calculated and compared by removing 5 numbers from the top and bottom. The results are shown in Table 1and Table 2.

Table 1 Descriptive statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Fraction	100	30.00	100.00	75.5200	9.64258
Valid N (listwise)	100				

Table 2 Trimmed mean descriptives tatistics

	N	Minimum	Maximum	Mean	Std. Deviation
Fraction2	90	63.00	89.00	75.8333	6.37807
Valid N (listwise)	90				

From the above figure, it can be seen that the tail-cut data is more concentrated and the variance is smaller, which is more suitable for data analysis. In application, not only students evaluate teachers' teaching, but also in other aspects of teaching evaluation, such as peer evaluation and teaching supervision data analysis, the theory of trimmed mean can also be used. Generally speaking, as long as the tail-cutting rate is well controlled, the obtained tail-cutting mean is a very good statistic, which can exclude the influence of extreme values, and can be widely used in teacher teaching evaluation[3-6].

2.2 Standardization of scores

Inconsistent measurement units, inconsistent scoring standards, and inconsistent scoring preferences will all have an impact on the results. For this reason, the method of standardization of scores can be used to solve this problem, that is, the score of each student is subtracted from the average score of the class and then divided by the standard deviation of the score of the class. The standardized data in a class will still be Large data is large, and small data is small, that is, the standardized data is order-preserving. For example, the scores of two students in a certain class are 70 and 90, the average of the class scores is 85, and the standard deviation is 10, then the scores of the two students after standardization are -1.5 and 0.5, which is still order-preserving. The standardized score is a non-dimensioned score with the average score of the class as a reference and the standard deviation as a comparison. According to statistical knowledge, when the sample is large, the standardized data will present a normal distribution with an expectation of 0 and a variance of 1. When comparing the scores of different majors and different classes, it is more necessary to use standard scores.

For example, in the evaluation of a teacher, class A scored 75 points, class A's average score was

70, and the standard deviation was 10; class B's classmate scored 80, and class B's average score was 75. Standard the difference is 10. From the data, 80 is greater than 75, but it cannot be said that class B has a higher score than class A. In fact, after the calculation of the standard score, the standard score of A is 0.5, and the standard score of B is also 0.5, which means that the two students AB are basically the same in the corresponding class. The grades given by a classmate to the teacher are actually quite comparable. Standard scores have many of the above-mentioned advantages. Using standard scores to make horizontal comparisons of data and compare the teaching evaluation scores in different classes and courses undertaken by the same teacher is comparable and is worth introducing and paying attention to in teaching evaluation. Thinking together with standardized scores and ordinary scores can analyze the relativity and absoluteness of data in some data, and can better analyze the real results of the data.

3. Data analysis methods –reduce the dimension of evaluation by big data analysis

In order to evaluate the teaching level of teachers, an appropriate index system should be established. For example, a university has designed an education evaluation index system, involving (1) Putonghua standards, and classroom arrangements are reasonable. (2) Get to and from get out of class on time. (3) The class is logically coherent and coherent. (4) Integrate theory with practice and do not follow the script. (5) The teaching content is moderate and substantial. (6) Pay attention to the frontiers of disciplines. (7) Appropriate amount of homework and careful correction. (8) Blackboard writing standards can effectively utilize modern teaching. (9) Pay attention to the cultivation of students' ability. (10) Focus on innovative thinking and curriculum ideology. 10 points for each item, out of 100 points. The system reflects the teaching characteristics of teachers in terms of teaching content, teaching attitude, teaching methods, and teaching effects. Currently, 20 teachers who have taught a certain class participate in the teaching evaluation. Now, a sample survey has been conducted on 50 students. Obtained 1000 questionnaire data, now the key is how to use the data to make analysis.

It is not good to calculate the average score of each teacher based on the total score given by each student to each teacher, which does not truly reflect the kernel results. Therefore, big data analysis methods can be used to perform statistical analysis on the teaching evaluation data. The traditional clustering method in cluster analysis can be used to draw a tree diagram of teachers' clustering, and then teachers can be divided into two groups according to the trend characteristics of the data. For some categories, K-means rapid clustering method or systematic clustering analysis method can also be used to divide teachers into the required number of categories.

It is easier to see the advantages and disadvantages of the same teacher category through cluster analysis, and it is more scientific and reasonable than the division of teacher categories based on simple scores.

Of course, the classification made by cluster analysis often only presents the final result. To clarify the content of these teachers' grades, it is necessary to perform principal component analysis or factor analysis on the data, and find out the principal components with a large classification contribution rate, or make use of the factor loading after factor rotation transformation, and classify some indexes as a certain factor. For example, the above indicators are divided into teaching content factor, teaching attitude factor, teaching method factor, and teaching effect factor.

Therefore, when doing teaching evaluation, we should design evaluation indicators scientifically and reasonably, present rich teaching evaluation categories, and then analyze the data kernel by means of multivariate statistical analysis to reveal the real evaluation results. A questionnaire survey was conducted on the evaluation index system listed above, and the clustering dendrogram was obtained after preprocessing the data, as shown in Figure 2.

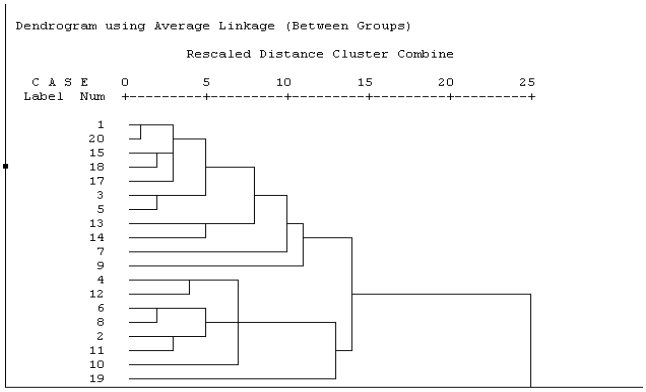

Figure 2 Dendrogram

It can be seen from the results in Figure 2 that if teachers are divided into 4 categories, teacher 16 is a category, teacher 19 is a category, teachers 4, 12, 6, 8, 2, 11, 10 are a category, and the rest of the teachers are a category kind. Only from the cluster analysis, only the classification of teachers can be seen, and the role played by the specific index system cannot be seen. Factor analysis can be done again.

As shown in Table 3, the KMO value is 0.567, which can be used for factor analysis. After factor analysis, it is concluded that 4 eigenvalues are greater than 1, and 4 factors can be analyzed, as shown in Table 4. Using factor rotation, the four factors are calculated as teaching content factor, teaching attitude factor, teaching method factor, and teaching effect factor, as shown in Table 5. Using factor analysis can reduce the dimension of data analysis and find convincing factors to explain the results of the assessment.

Table 3 KMO and Bartlett's test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.567
Bartlett's Test of Sphericity	Approx ChiSquare	99.115
	df	45
	Sig.	.000

Table 4 Total variance explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.910	39.100	39.100	3.910	39.100	39.100
2	1.974	19.739	58.839	1.974	19.739	58.839
3	1.173	11.728	70.567	1.173	11.728	70.567
4	1.019	10.191	80.758	1.019	10.191	80.758
5	.710	7.104	87.863			
6	.606	6.058	93.921			
7	.250	2.500	96.421			
8	.187	1.867	98.289			
9	.109	1.092	99.381			
10	.062	.619	100.000			

As shown in Table 3, the KMO value is 0.567, which can be used for factor analysis. After factor

analysis, it is concluded that 4 eigenvalues are greater than 1, and 4 factors can be analyzed, as shown in Table 4. Using factor rotation, the four factors are calculated as teaching content factor, teaching attitude factor, teaching method factor, and teaching effect factor, as shown in Table 5. Using factor analysis can reduce the dimension of data analysis and find convincing factors to explain the results of the assessment[7-8].

4. The source of the data is scientific

4.1 Consistency of peer evaluation and school supervision and evaluation data

Students are the direct participants in teachers' teaching activities, but they are not the only participants, and the diversity of participants in teaching assessment should be maintained. Peer evaluation and school-level teaching supervision evaluation should be added to the teaching evaluation of college teachers. Therefore, on the whole, the teaching evaluation of college teachers is at least a comprehensive analysis of three levels: students, peers, and supervisors. The scores between the three levels in the teacher evaluation are independent, and there should be no interaction items. And the scores of these three levels should have a strong positive correlation.

If a teacher gives a high evaluation in the teaching evaluation, peer or school-level supervision evaluation, for example, the average score given by 5 experts is 93. But student ratings gave teachers a low rating, with an average of 72 given by 50 students. Then the quality of the teacher's teaching is questionable. This is worthy of further study of the teacher's teaching level. In fact, only the consistency of the data in the three levels of teaching evaluation can often reflect the teaching level of the teacher more truly. Therefore, while doing a good job of students' teaching evaluation of teachers, the teaching supervision evaluation of peers or experts above the school level is also a means to prove the explanatory power of data. This multi-level evaluation model should be promoted in the teaching evaluation of college teachers.

Table 5 Rotated component matrix[a]

	Component			
	1	2	3	4
Putonghua standards, and classroom arrangements are reasonable.	.237	.816	-.097	-.018
Get to and from get out of class on time	.341	.797	-.013	-.332
The class is logically coherent and coherent.	.680	.412	.226	-.129
Integrate theory with practice and do not follow the script.	.255	-.726	-.259	-.116
The teaching content is moderate and substantial.	.837	.000	.213	.372
Pay attention to the frontiers of disciplines.	.129	-.090	.064	.918
Appropriate amount of homework and careful correction	.809	-.034	.313	-.192
Blackboard writing standards can effectively utilize modern teaching	.775	.185	.138	.252
Pay attention to the cultivation of students' ability.	.351	.080	.897	-.011
Focus on innovative thinking and curriculum ideology.	.209	.023	.924	.118

Here, we analyze the correlation between student evaluation scores and peer evaluation scores, and do hypothesis testing. The student's evaluation score is $x_1, x_2, \cdots, x_m$, the mean is $\bar{x}$, the variance is s_x^2 , and the student's overall expectation is μ_1. The rating of peer review is $y_1, y_2, \cdots, y_n$, the mean is $\bar{y}$, the variance is s_y^2, and the overall expectation of peers is μ_2 . The number of class members participating in the teaching evaluation and the number of peer reviewers are generally not large, so the approximate t-test of the two normal population means can be used here.

Null hypothesis $\mu_1 = \mu_2$,alternative hypothesis $\mu_1 \neq \mu_2$, $\alpha = 0.05$, Test statistics $t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{m} + \frac{s_y^2}{n}}}$, deny domain $\left\{|t| \geq t_{1-\frac{\alpha}{2}}(l)\right\}$, $l = \frac{s_0^4}{\frac{s_x^4}{(m-1)m^2} + \frac{s_y^4}{(n-1)n^2}}$, $s_0^2 = \frac{s_x^2}{m} + \frac{s_y^2}{n}$.After test score collection and data preprocessing, the following data are available $m = 50, \bar{x} = 81, s_x^2 = 25$, $n = 5, \bar{y} = 77, s_y^2 = 16$, after operation $t = 1.08, l = 4.33$, Integer 4, $t_{0.975}(4) = 2.78$, Obviously, the t value is not in the rejection domain, and the null hypothesis should be accepted, and it is believed that the student's score and the peer's score are consistent at this time. The consistency of the data does not mean that the data must be completely equal, it only needs to satisfy the hypothesis testing conclusion. Of course, if the number of data is large, other hypothesis testing methods can also be considered.

4.2 Separation of teaching and examination, focusing on the correlation analysis between students' performance and teachers' teaching quality

For the rationality of the test scores that students play, the separation of teaching and test should be achieved. On the other hand, students' scores belong to the subjective scores of teachers' teaching evaluation, which have great subjective initiative, and what we need is more objective and real data. Students' test scores can be a real and effective data that exists objectively. Why the scores of the college entrance examination can be used as a powerful reference factor in the evaluation of the teaching quality of a certain subject in high school, because the college entrance examination or other formal examinations in high school have achieved the separation of teaching and testing, and the data is quite objective and convincing.

There is a strong positive correlation between teachers' teaching quality and students' academic performance, but it is difficult to reflect it in the current mode of teaching and testing in many colleges and universities. Therefore, as long as the separation of teaching and examination in college teaching is done well, the teacher in charge of the course only talks about the content of the course. The correlation between grades and teachers' teaching quality can be more easily reflected. Even if different teachers teach the same course, the teaching evaluation based on students' scores should be more explanatory than the data that students directly rate the teachers. Moreover, in the classification of the same major in colleges and universities, there is no situation where students with better foundation are intentionally divided into a certain class. The excellent composition of students in different classes is balanced, which further ensures the correlation between the scores of students and teachers' teaching evaluation. For example, if the same teacher teaches 1 class of AB courses, before the separation of teaching and examination, 1 class A course takes 75, and after teaching and examination is separated, B course takes 80, then 80 points of B course are often more useful. A basis in this teacher's teaching evaluation, the data is more convincing to use.

In fact, not only the scores of students, but also in peer evaluation and teaching supervision evaluation should also pay attention to the relevance issues. The peer evaluators and teaching supervisors of the participating teachers should try to avoid being too familiar with the teachers involved, and should avoid favoritism for the sake of the data. Teaching quality data is more explanatory and more objective, and the school level can also consider this third-party peer evaluation or teaching supervision model.

5. Data reliability and validity analysis

In order to improve the reliability of the assessment, reliability analysis should be done, and a teacher's teaching reliability assessment should be conducted at the mid-term and at the end of a semester to see whether the re-test reliability score is higher. And in order to improve the reliability

of the data, the assessment should be done by a third party. The teachers who are tested should be kept confidential, and the final grades should be reported to the relevant teachers after the final results are announced, so as to avoid students being afraid of the authority of teachers' grades and not dare to make objective and fair grades. Perfunctory, answering questions hastily. For example, for the same course, mid-term and final, the same class gives the teacher's test scores an average of 75 and 95. The gap between the two is so big, then you should analyze the deep-seated reasons and find out What is the reason that the scores before and after the test are so inconsistent, otherwise it would be unreliable to directly use these two scores for educational evaluation.

It is also necessary to pay attention to the validity of teachers' teaching evaluation. The design of the questionnaire indicators should be scientific and reasonable, to avoid the emergence of systematic errors, to control the size of random errors, to pay attention to the degree of participation of students, and to have evaluation criteria for the setting of scores. Can't generalize. For example, evaluation of teachers' teaching materials, classroom teaching, and after-class reflection cannot be divided into equal proportions. Obviously, classroom teaching is a direct presentation of students' teaching and should occupy an important position. Only by designing teaching evaluation indicators and evaluation standards can the questionnaire be effective higher degree.

The distinction degree of the questionnaire should also be designed well, for example, in life, there are sometimes questionnaires with a perfect score of 10. When the number of participants is too large, the same score will be higher, which is not conducive to the understanding of the real situation of the data and the analysis of the data. Therefore, when designing the questionnaire, it should be at least a 100-point or 150-point system, and it is not allowed to set a starting score of 90 or more for the sake of the data. Only by collecting data fairly and objectively can the data be more effective and the real teaching evaluation results be analyzed.

6. Dynamic analysis of evaluation

Teaching evaluation is not the purpose of teaching, but only a means. Teaching evaluation is not to solve the problem of teaching evaluation once and for all, but to maintain a dynamic evaluation state. The teaching evaluation of some schools is only aimed at new young teachers, and only focuses on young teachers. New teachers do lack teaching experience. In contrast, there will be more problems in teaching evaluation. There are indeed areas for improvement. But teaching is a dynamic process. It does not mean that a course that you have been taking for 10 years is a good way to use the same method. The course must be a combination of cutting-edge disciplines and new things must be added. Therefore, teaching evaluation should also be a dynamic process. As long as teachers who have participated in front-line teaching should participate in teaching evaluation, they should reflect on and improve their own shortcomings, and strive for excellence in their areas or modules, so that they can take a good course. Give a good class, and give a gold class. Through the mathematical statistics and big data analysis methods mentioned above, the teaching evaluation model can be constructed scientifically and dynamically in order to do a good job in teaching evaluation.

Acknowledgments

This work was financially supported by Scientific research project of Sichuan Minzu College XYZB2013ZB、Teaching Reform Project of Sichuan Minzu College JG202147 fund.

References

[1] Yu Jiazhen, Zheng Yinxue, Wang Xinchun. Research on improving the teaching ability of vocational school teachers with "teaching evaluation" [J]. Vocational education,2019,12(41):46-50.

[2] Liu Zhentian. Classroom Teaching Revolution in Colleges and Universities: Actuality, Essence and Practice [J]. Higher Education Research, 2020, 41(7):58-69.
[3] Liu Xiumei, Gu Ming, Guo Hongting, Chen Shiyan. Research on the ethical dilemma of middle school teachers in teaching evaluation [J]. Chinese Journal of Education, 2022, (1):86-91.
[4] Jin Shengju. The disciplinary worries of big data education evaluation: A philosophical review of the instrumentalization of education [J]. Educational Research, 2019(8):33-41.
[5] Wang Xiaojie, Song Naiqing, Zhang Feiyi. Research on the evaluation index system of primary labor education--Based on the CIPP evaluation model [J]. Educational Research and Experiment, 2020(6):61-68.
[6] Xi Xiaoming. Educational assessment and artificial intelligence technology assessment in the field of learning [J] China Exam, 2121(5):56-62
[7] Zhang Pingping, Yuan Yuzhi. An analysis of the research on students who are resistant to adversity in the international education assessment project. [J] Comparative Education Research,2021(6):89-96.
[8] Peng Bo, Wang Weiqing, Zhang Jinliang, Yuan Jianlin, Yu Jianbo. How is education evaluation reform possible from the perspective of artificial intelligence [J]. Contemporary Education Forum,2021(6):1-15.