



Application of K-means Method Based on SPSS in Graphic Design Score Analysis

Jing Luo^(✉)

Guangzhou Nanyang Polytechnic College, Guangzhou, China

85810733@qq.com

Abstract. Usually the method of personnel training examination cannot fully verify the effect of teaching reform. Therefore, SPSS statistical software was used for α reliability coefficient analysis and K-means cluster analysis. In this paper, K-means algorithm is used to analyze and classify students' graphic design scores, discover the distribution characteristics of students' scores, understand the effect of "hybrid + ideological and political" teaching reform, provide a basis for teaching reform and improvement, and accurate reform of teaching models and methods.

Keywords: α Reliability Coefficient · K-means · SPSS

1 Introduction

In November 2020, the second National MOOC Education Innovation Conference and University Online Open Courses Joint Conference was held in Beijing. In the conference, an important result was released – the Guide to the Construction and Application of MOOC in colleges and universities [1]. The guideline defines the construction objectives, builds large-scale online open courses, improves the quality of talent training and speeds up the modernization of higher education [2].

From the perspective of higher vocational engineering course teaching, the teaching mode is gradually developing from offline to blended teaching, and the teaching content is integrated with ideological and political elements. But the ideological and political elements of the course only stay on the surface, most courses attach a single ideological and political element. The ideological and political elements of engineering courses have not been thoroughly explored, especially the flexible integration of ideological and political elements into the teaching methods of information technology at the present stage, which is the difficulty in the construction of computer courses.

2 Research Methods and Innovations

The "hybrid + ideological and political" teaching model is established to carry out teaching experiments in graphic design courses. After the completion of the mid-term

teaching, the project examination will be conducted. By analyzing students' project achievements, the "hybrid + ideological and political" teaching reform is promoted [3].

The reliability of the questionnaire was tested by α reliability coefficient, and students were asked to evaluate knowledge ability and personal literacy independently. K-means algorithm was used to carry out cluster analysis, observe the distribution of each attribute, and timely improve the "hybrid + ideological and political" teaching content.

3 Theoretical Basis

3.1 SPSS Analysis

Spss analysis consists of the following technical links: use the data management module to establish data information, data interface is similar to table, relatively general. Then, the variable module is given reasonable variable information. Finally, a simple visual interface is used to complete various data algorithms [4].

3.2 K-means Clustering Definition

Cluster analysis is to find the relationship between data objects in the data and group the data. The greater the similarity within the group, the greater the difference between the groups, the better the clustering effect.

K-means clustering algorithm is a cluster analysis algorithm with iterative solution. It randomly selects K objects as the initial cluster center, then calculates the distance between each object and each seed cluster center, and assigns each object to the nearest cluster center. The cluster centers and the objects assigned to them represent a cluster. Each time a sample is assigned, the cluster center of the cluster will be recalculated according to the existing objects in the cluster. This process is repeated until some termination condition is met. The termination conditions can be that no object is reassigned to different clusters, no cluster center changes again, and the sum of squared errors is locally minimum [5].

3.3 K-mean Clustering Step

- Firstly, determine the value of k, and obtain k sets through clustering. ($1 \leq i \leq k$) Depending on the size of the class, 3 to 6 sets can be selected appropriately.
- Select k, data points randomly from the data set as the center of mass.
- For each point in the data set, the Euclidean distance from each centroid c_i is calculated. If it is close to the centroid, it will be divided into the set to which that centroid belongs [7].
- All data are classified, and there are a total of k sets. And then recalculate the center of mass of each set.
- If the distance between the newly calculated centroid and the original centroid is less than a certain threshold (indicating that the position of the recalculated centroid does not change much, tends to be stable or converges), we can consider that the clustering has reached the desired result and the algorithm is terminated [8].

The specific formula of the algorithm is:

$$SSE = \sum_{i=1}^k \sum_{x \in c_i} dist(c_i, x) \tag{1}$$

k in Eq. 1 represents that the dataset is divided into k clusters, c_i represents the i th cluster center, and $dist$ represents the Euclidean distance between two points.

4 Models and Data

4.1 The Establishment of “Hybrid + Ideological and Political” Teaching Model

The graphic design course is divided into basic project, breakthrough project and innovation project. These projects are integrated with ideological and political elements from the acquisition of materials, the formulation of project objectives, and the grasp of details. The specific “hybrid + ideological and political” teaching model is shown in Fig. 1.

4.2 The Data Collection

Graphic design course is a basic professional course of computer major, which involves many classes and the opening time of different majors is different. In this study, 20 series media majors are used as experimental samples.

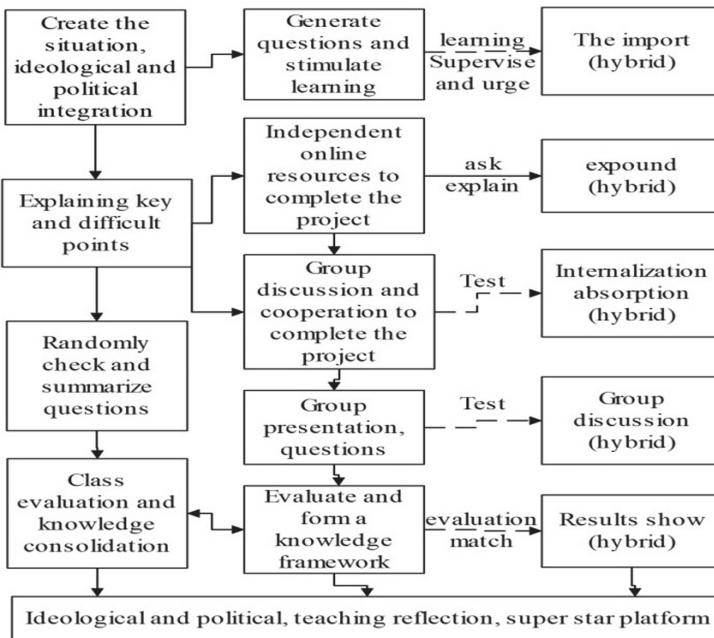


Fig. 1. Design of “Hybrid + Ideological and Political” Teaching Model

cluster	jbcb	xq	hb	tc	wz	db
1	100.00	95.00	86	86	95	100
2	85.00	96.00	86	86	95	98
3	95.00	97.00	86	86	95	85
4	100.00	98.00	98	98	98	98
5	95.00	99.00	86	86	95	83
6	95.00	100.00	86	86	95	85
7	100.00	98.00	98	98	98	98
8	95.00	95.00	95	95	95	83
9	75.00	95.00	95	95	95	85
10	95.00	95.00	95	95	95	98
11	95.00	95.00	95	95	95	85
12	95.00	95.00	95	95	95	98

Fig. 2. Assessment Data of some Students in Class 2

After the completion of the middle study period, the project examination will be carried out. The project examination is divided into 6 assessment points, which are basic operation, selection, paintbrush, layer, text and defense. The first five knowledge points focus on professional skills and innovation ability, while the last one examines personal literacy. Class 1 adopts traditional teaching, and class 2 adopts “hybrid + ideological and political” teaching. The final score of Class 2 is better than class 1 on the whole, so the score of Class 2 is used as the data set. The specific test score is shown in Fig. 2.

4.3 The Research Methods

SPSS software was used for statistical analysis of the experimental data, α reliability coefficient was used to confirm the credibility of the questionnaire survey, and K-mean clustering was used to analyze the results of “hybrid + ideological and political” teaching.

5 Comparative Study Results and Data Analysis

5.1 The α Reliability Coefficient Method Was Used to Test the Reliability of the Questionnaire

A total of 92 people participated in the test, with 92 questionnaires and 92 valid questionnaires. In order to test the reliability of the questionnaire, the reliability coefficient of the questionnaire data was calculated. The number of questions in the questionnaire was 15. The α reliability coefficient method was used, and the formula was as follows:

$$\alpha = \frac{k}{k - 1} \left(1 - \frac{\sum_{i=1}^k s_i^2}{s_x^2} \right) \tag{2}$$

The k value in Eq. 2 is the total number of questions in the questionnaire, s_i is the variance of the score of question i , s_x is the variance of the total score of all questions. Through calculation, the reliability coefficient value is 0.825, the data reliability is good, the data can be referred to.

Table 1. Clustering Results of $k = 3$

K-mean Cluster Analysis	Clustering		
	1	2	3
Basic operation	96.21	82.00	93.31
District	96.11	91.43	87.38
The brush	92.58	84.57	79.92
The layer	88.79	92.43	79.00
The text	94.89	91.43	93.69
Plea	93.42	72.86	87.23
Number	19	14	13

5.2 Analysis of Clustering Results

In order to verify the rationality of carrying out “hybrid + ideological and political” teaching resources in the graphic design course and to carry out student-centered teaching, students with similar levels were grouped according to the course assessment requirements, and teaching resources were increased or decreased according to the needs of different groups. Teachers can also select students to participate in competitions and complete school-enterprise cooperation projects according to students’ knowledge and ability, so as to cultivate students’ skills and humanistic qualities more pertinently.

There are 46 students participating in the clustering, which is too many groups to show interval, and too few groups to find excellent students. Therefore, the K value is set as 3 and 4 to carry out scientific grouping. The clustering results are shown in Table 1 and Table 2.

As can be seen from Table 1, the current iteration is 3, and the minimum distance between the initial centers is 38.678. The first team is excellent as a whole and has both professional skills and humanistic qualities. There are 19 members in total who can participate in the competition. The second group, the course ideological and political education effect is not good, professional skills, a total of 14 people, need to add humanistic care, ideological and political education in the subsequent teaching. The third group has two knowledge weaknesses, and needs to add learning videos and exercises. There are 13 people in total. Through these data, it is found that the current ideological and political elements need to be further mined and the points need to be smaller [6].

As can be seen from Table 2, the current iteration is 4, and the minimum distance between the initial centers is 29.665. The fourth team is excellent as a whole, with professional skills and humanistic qualities. There are 11 members in total who can participate in enterprise projects. The first group had a general knowledge of layers, a total of 20 people. The second team, 9 people in total, has a poor grasp of ideological and political education and constituency knowledge points, and needs to add learning videos and exercises. The third group has a general knowledge of painting brush and ideological and political education, and needs to add learning videos and exercises. There are 6 members in total.

Table 2. Clustering Results of K = 4

K-mean Cluster Analysis	Clustering			
	1	2	3	4
Basic operation	93.00	85.00	83.00	100.00
District	96.00	73.00	95.00	98.00
The brush	95.00	89.00	65.00	98.00
The layer	75.00	90.00	90.00	98.00
The text	90.00	93.00	90.00	98.00
Plea	83.00	75.00	70.00	98.00
Number	20	9	6	11

Through the analysis of the clustering results, the “hybrid + ideological and political” teaching resources and teaching model are reasonable and effective, and the case resources need to be increased in the future.

6 Conclusion

Cluster analysis is a common method in machine learning, which can quickly and efficiently group data sets, and has been applied to various industries. In this paper, the reliability analysis and the classification of students’ scores are completed by SPSS. Through the K-means algorithm in cluster analysis, students majoring in logarithmic media carry out the reform of “hybrid + ideological and political” teaching mode, and carry out experiments in the course of graphic design. The feasibility of cluster analysis is verified by testing 46 sample data. This also provides a basis for teachers to adjust the teaching resources of the super star platform, improve the teaching design, and scientifically and reasonably explore the ideological and political elements.

The K-means algorithm is used to dig the data deeply, which is convenient for us to find excellent students, promote the teaching reform in time, and improve the teaching quality. The combination of k-means algorithm and teaching reform is very meaningful, but it cannot be verified in a unilateral way, and other methods need to be explored for further research.

Project Fund. 2021 University-level Innovation and Strong School Project “Teaching Reform and Practice of Ideological and Political Curriculum Implementation in Blended Teaching of Graphic Design under the Concept of OBE Education” (NY-2021CQ-JGYB003). Research on Ideological and Political Practice Path of Higher Vocational College Curriculum from the Perspective of “Three Integrity Education” (NY-2021CQ-JGYB002). Study on the cultivation and practice of students’ higher-order thinking ability under mixed teaching mode (NY-2021CQ-JGYB004).

References

1. Li Shangrong. An empirical study on the Influence of teacher-student Co-construction of Online Courses on learning Autonomy [J]. Heilongjiang Education (Theory and Practice), 2022(09): 66-68.
2. Xiao Xiaoxu. Practical Research on the Hybrid teaching Mode of “MOOC + SPOC + Flipped Classroom”– A Case Study of “Python Basics” course [J]. Information and computer (theoretical edition), 2020, 32(13): 244–245.
3. Xu Lei, Zhang Genshan, Cui Jian. Exploration and Practice of Blended online and offline teaching Model: A Case study of “Modern Educational Technology” course [J]. Journal of Shijiazhuang University, 202, 24(04): 128–132.
4. Qi Feng, Yang Yichen. Skillfully using SPSS cluster analysis to predict outliers [J]. China Metrology, 2022(05): 89-91.
5. Wang Ting. Online Learning behavior path analysis and Research based on K-means clustering Algorithm in the background of big data [J]. Network Security Technology and Application, 2022(05): 43-45.
6. Luo Jing, Research on the cultivation path of “Craftsman Spirit” for high-quality technical and skilled talents under the Modern Apprenticeship System [J]. Education in Guangdong. 2020(04): 4-8
7. Lin X Q. Research on online learning behavior based on K-means clustering algorithm in the era of big data [J]. Electronic design engineering, 2021, 29(18): 181-184+193.
8. Lyu D. Study on student performance data analysis and prediction Modeling based on K-means clustering algorithm [J]. Microcomputer applications, 2021, 37(05): 148-150.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

