

Research on the Application of R Software in the Teaching Analysis of University Computer I

 Zhang Qi¹ Yang Atiao¹

¹School of Big Data and Computer Science, Guizhou Normal University
Guiyang City, Guizhou Province 550001

 Yang Yuan²

²School of International Tourism and Culture, Guizhou Normal University
Guiyang City, Guizhou Province 550001

Abstract—In order to improve the teaching effect of University Computer I course and promote the cultivation of students' computational thinking, this article uses Rattle package of R software to perform data analysis on the end-of-term computer examination score for 2015 Grade science specialty in University Computer I at Guizhou Normal University. The analysis mainly includes three aspects: the analysis of all score data samples, the analysis of the impact degree of the professional course scores, and the analysis of the influence degree of the course scores set by the question type. The analysis result shows that, the main core factors influencing the course learning effect of University Computer I include the difference of students' learning capacity and the setting of the course content. By focusing on these two core factors, this paper puts forward a teaching reform approach to build course learning platform to enrich course learning resources, strengthen course operation and application training, set up professional elective modules, and reform the model of teaching evaluation.

Keywords—University Computer I; R Software; Teaching reform

I. INTRODUCTION

University Computer I is the general knowledge required course for undergraduate students majoring in non-computer specialties, and the main contents of the course include computer system overview, the basic operation of windows operation system, word text processing, the basic operation and application of Excel and PowerPoint, as well as the basic concept and operation of computer network. On the basis of cultivating students to master the basic principle of computer and the basic knowledge, through the practical operation of computer, it is aimed to improve the students' application capacity of operation system, cultivate the basic skills of information processing, and improve the capacity of computational thinking mode.

Educational Data Mining (EDM for short) [1-3] conducts mining and analysis of education related data through comprehensively applying pedagogy, statistics, machine study, data mining technology and method, and these data will be converted to information that can be utilized by teachers, students and other education related personnel. As an open-sourcing statistic and data mining software, R software covers from statistic calculation to machine study, from financial analysis to biological information, from social network analysis to natural language processing, and from various database and various language interfaces to high-performance calculation model. Due to cross-platform, and strong statistic analysis capacity and data visualization function, it is currently a relatively popular statistic and data mining tool [4].

This research utilizes Rattle package in R software to analyze University Computer I online computer test scores of totally 1,027 students from 26 specialties in 2015 Grade of Guizhou Normal University. Analyze the relevant factors influencing the course learning effect from the original discrete data, and then assist teachers to pointedly formulate reasonable teaching contents, improve teaching method, and promote teaching level and teaching effect, and further promote the cultivation of students' computational thinking.

II. METHODS OVERVIEW

Rattle is the graphic interaction interface of R used for data mining; through utilizing Rattle, it can fast solve common data mining issues. Rattle utilizes tab bar, which is as shown in Figure 1; it respectively ranks the six steps for complete data mining (comprehension questions, comprehension data, preparation data, modeling, and model evaluation and application) [1] from left to right, and then respectively complete the relevant step for data mining. Rattle package also provide many statistical magnitude, data collection and follow-up model evaluation. As a visual tool, Rattle can also create some very basic area chart, line chart, box plot, histogram, etc. Based on the good interaction of Rattle and R platform, the research applies Rattle package to conduct data analysis on the online computer test of University Computer I.

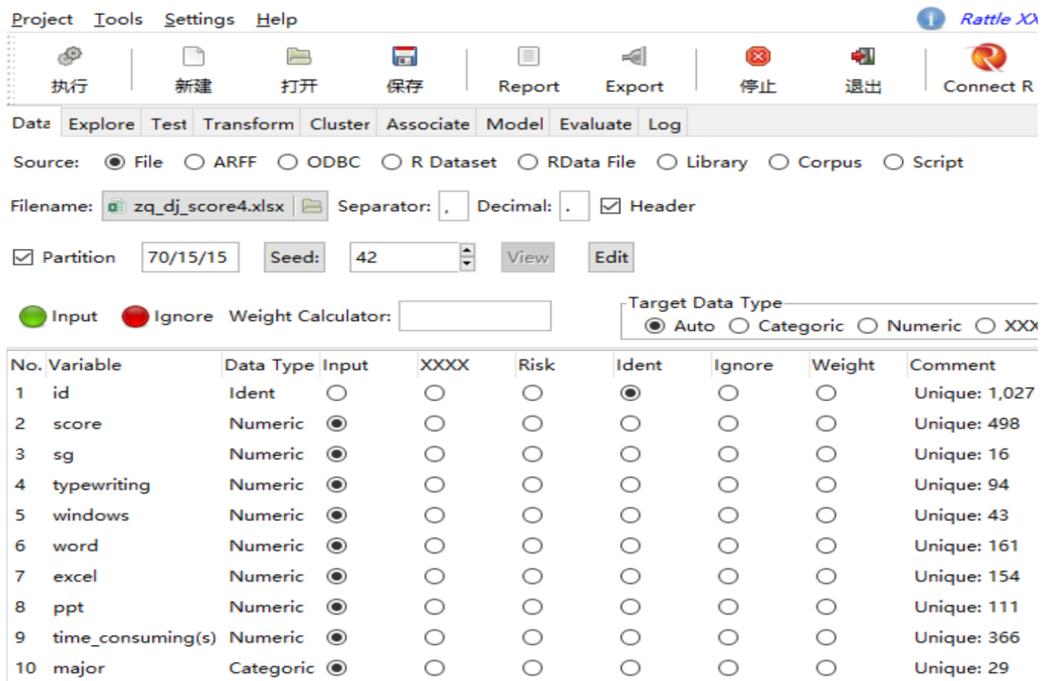


Fig. 1. Rattle Working Interface

III. ANALYSIS ON THE STUDENT SCORE OF UNIVERSITY COMPUTER I

A. Analysis on the overall data samples

Conduct statistical analysis for all term-end scores of University Computer I and make histogram, which is as shown in Figure 2. According to the result, in 1,027 sample data, the average score is 57.53544 points, the highest score is 91.2

points, and the lowest score is 10.8 points, the standard deviation is 16.19598, the sample variance is 262.3099, the asymmetry coefficient is -0.3080715, and the kurtosis coefficient is -0.4218509. In combination with the term-end score statistic histogram and QQ (Quantile-Quantile Plots) Figure, which is as shown in Figure 3, the quantity of people with the score less than 60 points is more than that with the score above 60 points, and the difference is big; besides, it can be seen that, it cannot meet the normal distribution.

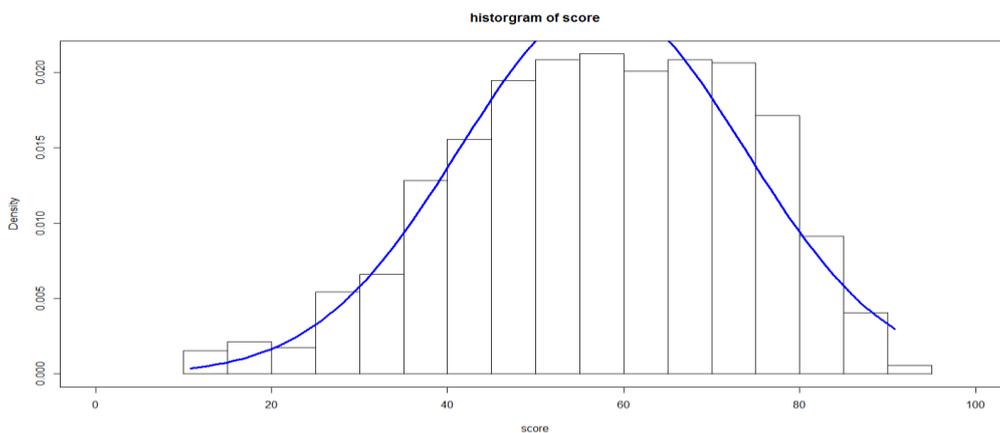


Fig. 2. End-of-term Score Histogram

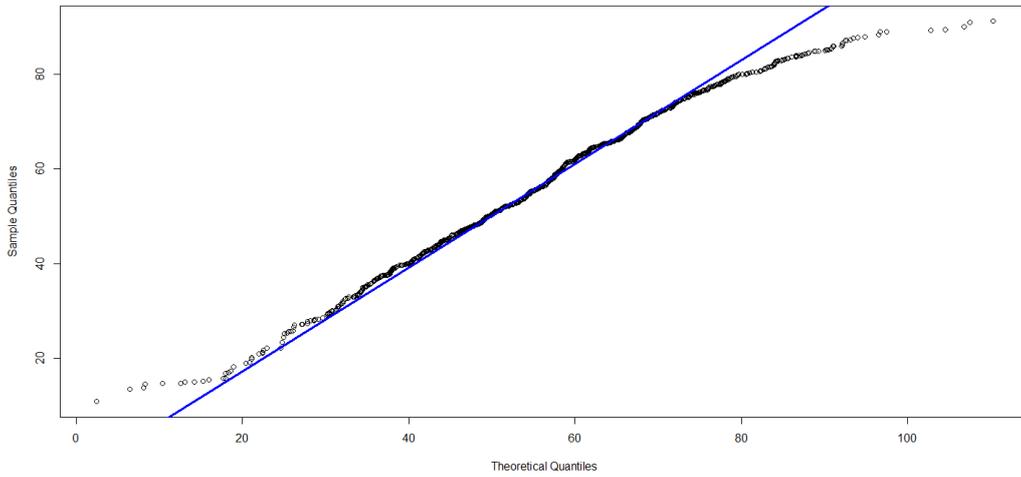


Fig. 3. End-of-term Score Quantile-Quantile Plots Figure

B. Analysis of the influence of specialty on the course score

In order to discuss the influence degree of specialties on the score of University Computer I, conduct single factor variance analysis and Turkey multiple comparisons for university computer scores as per specialties [6]. The sample data involve 26 specialties; firstly, conduct single factor variance analysis on the students score data of various specialties, and the result

is as shown in Table 1. Through combining Table 1 with Figure 4, it can be seen that specialties can exert an obvious influence on the term-end scores of University Computer I. In order to further understand the influence of various specialties on the score, conduct clustering analysis on the average score of 26 specialties as per ward method, and the result is as shown in Figure 5.

TABLE I. VARIANCE ANALYSIS ON THE SCORE INFLUENCE OF UNIVERSITY COMPUTER I FOR VARIOUS SPECIALTIES

Data Source	Degree of Freedom	Quadratic Sum	Mean Sum of Square	F-ratio	P Value
Specialty	25	33836	1353.4	5.758	<2e-16
Error	1001	235294	233.3		
Total Sum	1026	269005			

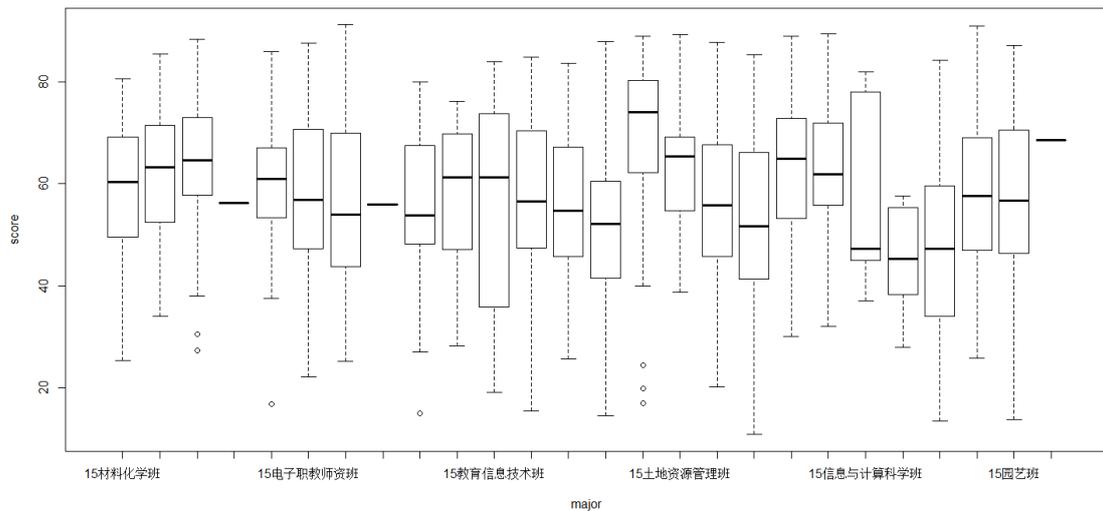


Fig. 4. Difference of Various Specialties

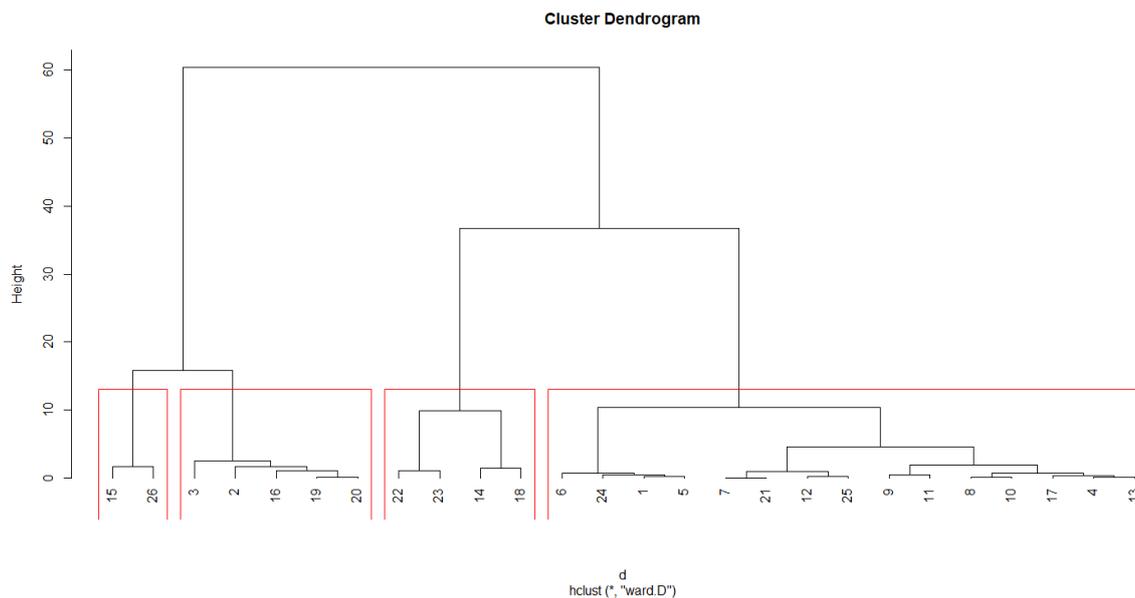


Fig. 5. Clustering Figure for University Computer I Score of Various Specialties

As per Figure 5, divide the score of 26 specialties into four types: the specialty with the highest score; similar score above 60 points; similar score below 60 points, but higher than 55 points; lower score, and in combination with the college entrance examination enrollment mark of Guizhou Normal University in 2015[7], it can be seen that students majoring in specialties with the highest college examination enrollment score (such as mathematics and applied mathematics major) can still take lead in study after entering the school, but students majoring in specialties with lower enrollment score in those years still fail to obtain ideal score after entering the school. The conclusion is that: different qualities of students can have certain influence on the teaching effect of University Computer I.

C. Analysis the the influence of question type setup on the course score

The term-end computer examination of University Computer I is constituted by 6 parts, including single selection, typing, windows, word, excel, and PowerPoint. This article selects Apriori [8] algorithm to analyze the hidden internal relation of different question types, and it is aimed to form a comprehensive and systematic computer course teaching contents correlation system, and provide guidance for the university computer course teaching management and students' learning strategies. In order to simplify processing, we conduct discretization for original data through editing basic data set, and if the total score is discretized to 90-100 as per the scores, it is A1; if it is 80-90, it is A2; if it is 70-79, it is A3; if it is 60-69, it is A4; if it is 60 points, it is A5, similarly, we conduct similar discretization for 6 question types of single selection, typing, windows, word, excel, and PowerPoint, and obtain the information sheet, which is as shown in Table 2.

TABLE II. INFORMATION SHEET OF DATA AFTER PRE-PROCESSING AND DISCRETIZATION

id	Total Score	multiple choice	Typing	Windows	Word	Excel	PowerPoint
1	A2	B4	C1	D2	E1	F1	G4
2	A2	B4	C1	D5	E1	F1	G1
3	A2	B3	C1	D2	E4	F2	G1
4	A2	B5	C1	D4	E1	F1	G1
.....							

The data analysis process adopts arules package provided by R language for completion, and takes apriori algorithm as the basis, sets the minimum support as 0.3, the confidence

coefficient as 0.8 for generative rules, and association rules grouping, which is as shown in Figure 6.

Grouped Matrix for 21 Rules

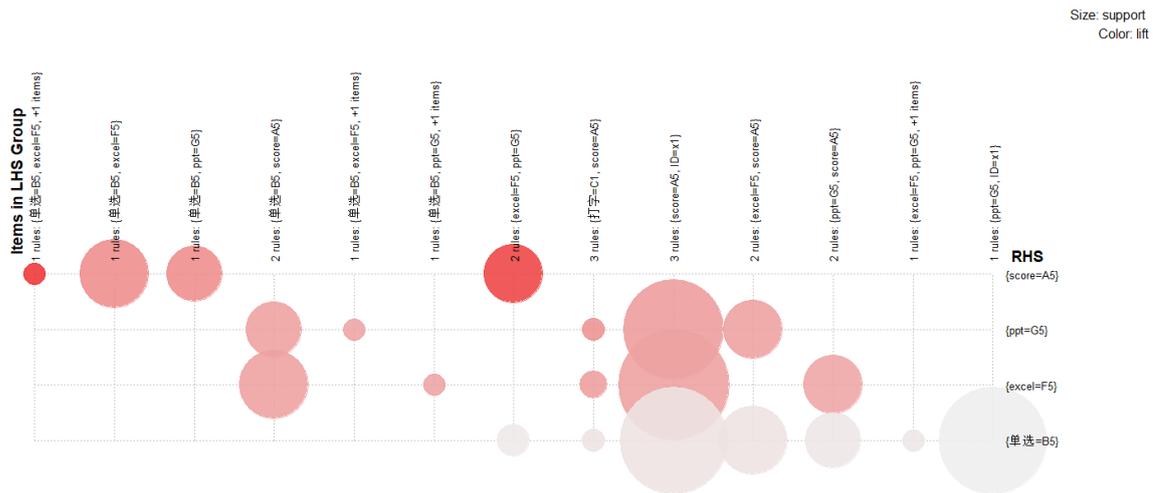


Fig. 6. Association Rules Grouping Figure

As can be seen from Figure 6, there are only two question types among the 6 ones possessing worse score, and the total score is worse too; as long as the question type with 20 point values acquires a bad score, the total score will be bad, the timely score has certain relations with the occupied score proportion. The conclusion is that: it is requested to pointedly cultivate students' windows, word, excel, and PowerPoint application capacity.

IV. CONCLUSION

Through the aforementioned analysis, there are mainly two core factors influencing the course study of University Computer I, one is the difference of students' learning capacity, and the other one is the factor of course contents setup. In order to solve these two core factors, the following teaching reform thought can be adopted: 1) Construct the digital course learning platform of University Computer I, provide rich learning resources on the platform, and pay attention to guide and cultivate students to form the habit of using digital resources to carry out autonomic learning, so as to meet the learning demands for different layers of students. 2) Reinforce the basic operation and application training of windows and office, additionally set professional elective modules in close combination with specialties, introduce various evaluation modes, and make the evaluation more scientific.

Besides, in the course teaching process of University Computer I, and the process for teachers, students and teaching manager to use various learning management and digital course learning platform, lots of learner information, learning process information and other data will be generated, and there is certain restriction for merely using the computer examination score data of University Computer I course to carry out data mining. With respect to this problem, we plan to integrate education data mining technology into the teaching process of University Computer I, and apply system design method to design the comprehensive platform for the teaching evaluation

of University Computer I course at Guizhou Normal University, so as to fully ply the data value in the learning process, convert these data into information that can be utilized by teachers, students and education related personnel, and meanwhile, provide support service for teaching decision-making, the improvement of teaching contents, the teaching sequence, teaching method, and teaching optimization.

ACKNOWLEDGMENT

Fund project: Supported by Guizhou Provincial Undergraduate Teaching Engineering Construction Project ([2016] (2) No. 17 University Computer I Digital Course Construction Based on Flipped Class Teaching Mode).

REFERENCES

- [1] Goyal M, Vohra R. Applications of Data Mining in Higher Education [J]. International Journal of Computer Science Issues, 2012, 9 (2).
- [2] Huebner R A. A Survey of Educational Data-Mining Research. [J]. Research in Higher Education Journal, 2013, 19:13.
- [3] Zhou Qing, Mou Chao, and Yang Dan. Education Data Digging Research Progress Summary [J]. Journal of Software, 2015, 26 (11):3026-3042.
- [4] Luis Torgo. Data Mining and R Language [M]. China Machine Press, 2013.
- [5] Bra P D. Knowledge Discovery with Genetic Programming for Providing Feedback to Courseware Authors [M]. Kluwer Academic Publishers, 2005.
- [6] Tang Yincai. R Language and Statistical Analysis [M]. Higher Education Press, 2008.
- [7] College Entrance Examination Enrollment Mark for All Specialties in Guizhou Normal University in 2015, [EB/OL] http://gkcx.eol.cn/schoolhtm/specialty/514/10035/specialtyScoreDetail_2015_10026.htm
- [8] Yang Chunguang. Analysis on Student Scores Based on Association Rules [J]. Computer Knowledge and Technology, 2017, 13 (20).