Journal of International Education and Development
2022, VOL. 6, NO. 9, 81-87
DOI: 10.47297/wspiedWSP2516-250015.20220609

The Curriculum Reform and Practice of "Sampling Survey" with the Core of Cultivating Students' Autonomous Learning Ability

Lei Wen, Xinyue Ma, Rufei Zhang

School of Economics, Hebei GEO University, Shijiazhuang 050031, China

ABSTRACT
The cultivation of college students' autonomous learning ability needs to go deep into each professional course. As the core course of statistical majors, "Sampling Survey" requires teachers to attach equal importance to theory and practice, which requires teachers to consciously link theory teaching with solving practical problems in classroom teaching. This paper provides a detailed analysis of the importance of R software for the course of "Sampling survey", and advocates that it should be introduced into the classroom auxiliary teaching of sampling survey based on many characteristics of R language. At the same time, combined with the front-line teaching practice, this paper puts forward strategies to optimize classroom design in terms of the specific arrangement of online and offline teaching content, the design of computer content and the assessment method, in order to improve students' autonomous learning ability by optimizing the existing sampling survey teaching mode.

KEYWORDS
Sampling Survey; R Langage; Self-learning Ability Improvement

1 Introduction

The cultivation of students' autonomous learning ability is inseparable from the combination of professional knowledge and software operation. Sampling survey is the core course of statistics major, which is followed by basic courses of mathematical statistics and other professional courses of social survey methods. The course is based on simple random sampling and is developed around sampling organization forms such as stratified sampling and cluster sampling. Because of the practical nature of sampling activities, the usual course organization is theory lecture-exercise. Any sampling survey book will also introduce many exercises for each sampling organization form to help students straighten out the relationship between various sampling organizations. At the end of the course, students are required to conduct simple sampling design and evaluate the quality of the obtained samples. Both sampling theory and application are included in the examination. Recently, more and more teachers have realized that such curriculum organization cannot support the demand of modern statistical work for talents, and the case in the form of word problems is no longer enough to cultivate students' sampling ability and data literacy. The introduction of software assisted sampling and estimation has become a trend.

2 Literature Review

In recent years, the reform of "Sampling Survey" course mainly focuses on improving students' practical operation ability through actual combat. For example, Li Hao advocated that students

should choose software and teachers should act as consultants [1]; Zhang Qinghong summarized his reform experience of Sampling Survey course over the years [2]and gave a detailed description of the sampling process; Li Chunping [3] and Hou Zhenmin [4] discussed the practice teaching process of the sampling survey course, and the former also designed the practice teaching plan from many aspects. More and more teachers, such as Huang Lifang [5] have introduced the use of software into curriculum reform. Zhang Xuexin [6] proposed to use SPSS and MATLAB software to learn sampling theory and explain the operation steps. Zhu Song [7] and Lu Yugui [8] made a summary of the experience of using SPSS to assist teaching. Any software can be used as an aid in college classes. R language is a programming software known for its open source, statistical computing ability and graphing ability. At present, with the continuous development of R language, it has been competent for high-end functions including data mining, database management, deep learning and so on. It is growing into a comprehensive data management tool, and it is increasingly common for teachers to choose R language to assist teaching. Li Xiongying [9] and Chen Zhigang et al. [10] advocated that R language system should be included in the study of Statistics, a basic course of economics and management, and elaborated in detail through cases. In recent years, R has been used in more than just statistics classes. Lv Xiaokang [11] especially combined R language with the characteristics of psychology teaching and put forward a new concept of psychology teaching relying on modern statistical tools. Li Hongzheng et al. [12] elaborated the method of introducing R software into financial time series. As R language has become the mainstream software in the field of statistical research, studies on the combination of R and sampling courses have gradually emerged. Zhao Peixin et al. [13] gave R language program without putting back random sampling. Wang Wei [14] and Lu Yugui [15] demonstrated the realization of different sampling organization forms in R language with complete cases. Throughout previous studies, it has become a consensus to introduce software teaching in sampling survey classes, while teachers' exploration focuses on knowledge points. At present, there are few sampling survey studies based on improving students' autonomous learning ability. This paper advocates the application of R language, an open source and professional software, to the sampling survey course, in order to further expand the auxiliary teaching and the cultivation of students' autonomous learning ability. All these need to be achieved through the optimization of teaching links.

3 A sample Survey on the Necessity of Introducing R Language into the Classroom

The necessity of the combination of sampling classroom and R language is reflected in the continuity of software learning, the characteristics of software itself, and the cultivation of students.

(1) Necessary for longitudinal use of software

The development of computer technology has injected vitality into the research and study of statistics. The statistics classroom "combined with software" not only reflects the application of statistics major, but also is an urgent need to cultivate practical talents. The students majoring in economic statistics in our school learned "The Basis of R Language Statistics" in the sophomore year, so it is reasonable to introduce R language operation into the sampling survey course. The combination of the sampling survey course and R language can not only realize the practicability of the course teaching, but also help to promote the extended use of R language for students.

(2) R software comes with rich and real data

Sampling technology can deduce the whole population through a single sampling. This feature

of "knowing the whole leopard through a glimpse" reflects the "black box" nature of the sampling process. Therefore, the teaching content of teachers mainly focuses on sampling estimation, especially on the calculation of the accuracy of estimation. However, for students, the process of sample acquisition is a necessary data processing ability and the premise of estimation accuracy. In teaching, the sample can be obtained through the actual survey to obtain the primary data, also can be obtained by generating pseudorandom numbers. As the "practice" data, the rich built-in data set in R language can be used conveniently for students, and these real data can expand students' learning ability and vision.

(3) R's powerful graphing capabilities extend ideas for data visualization

In the era of big data, the status of descriptive statistics is soaring. As a professional ability of students majoring in statistics, data visualization has been paid more and more attention by various schools. As with all inferential statistics, the first step in sampling estimation is to describe the current situation, and graphs are a powerful tool for descriptive statistics. The excellent mapping function of R language has won it a place in the era of big data. In terms of sampling survey methods, for example, stratified sampling and cluster sampling, all indicators of each layer or group need to be shown and plotted separately, which cannot be drawn in Excel that students are familiar with, while the faceted function in R can be easily qualified.

(4) Beneficial to students' data literacy

"The key to the cultivation of college students' data literacy lies in the cultivation of data skills. The cultivation of data skills should start from theory and practice, in order to train students to master the scientific principles of data processing, master the basic tools of data mining and data analysis, and have the ability to use data to serve life and work." [16] There is no doubt that the cultivation of data literacy of economic statistics major needs to be realized by software. R software is a complete set of data processing, calculation and mapping software system. A complete sampling estimation process involves : sample acquisition-data cleaning-descriptive statistics-interval estimation-report presentation and many other links, which can be completed independently by R.

(5) A powerful tool for students to improve their autonomous learning ability

Littk Wood considers "autonomous learning ability" as "the ability to think critically, make decisions, and perform independent actions independently" [17]. Unlike excel and spss, R's code manipulation makes it easy to collaborate between multiple people and different tasks, thus being a good boost to students' self-learning skills. Sampling survey is a core course of statistics majors. Students' autonomous learning ability can be summarized into three aspects: using software to solve existing problems, exploring new sampling methods and learning new software. Through the specialized package, R can easily realize the whole process from data generation to report presentation; The R language extension pack is changing rapidly, which makes it possible for students to use new methods of estimation; Students need to learn and explore how the new package is used. It is an important way to cultivate students' autonomous learning ability to understand new methods and solve sampling and estimation problems with all kinds of old and new methods by using R language.

(6) New liberal arts building needs R language support

Strategic, innovative and open are the key elements of the new liberal arts [18], and the free, open-source and all-purpose nature of the R language is a strong support for the construction of the new liberal arts. The author believes that the codification of statistical processing of R language is conducive to the teachers and students of liberal arts. The data visualization content can be used as a breakthrough point, and the R software can be introduced into broader liberal arts teaching and scientific research, which is of great benefit to the cultivation of data literacy of teachers and students.

The introduction of R language into the sample survey course is not only an important way for teachers to realize computer-aided instruction, but also can cultivate students' autonomous learning ability. By moving part of the course content online and realizing the organic interaction between online and offline, it is helpful to optimize the class time and make students develop the habit of independent learning. This requires efforts to optimize classroom arrangements and assessment methods.

4 A Sample Survey Curriculum Reform Approach Based on the Improvement of Students' Independent Learning Skills

(1) Online and offline parallel theoretical lectures

What is the proportion of online teaching and what is the content? This is a question that deserves constant consideration and adjustment. For the content of the combination of theory and practice, it is obvious to put the operation content in the network. Therefore, the time of theory teaching also needs to be optimized.

1) Classroom lectures: add and subtract both

It is a trend that the total class time in today's universities will be compressed, and so will the combination of online and offline classes. Under these two trends, what to say in class and how to say it has become the first question for teachers to think about. Sampling survey course content of the trunk is undoubtedly a few big sample organization form, will be lectures focus on the trunk, the little relationship with the main parts, such as questionnaire design, sampling different organizations use and non probability sampling, the sampling error given to students to use network resources self-study class, this is the course teaching of "subtraction" reflect.

Sampling theory is based on probability theory and the law of large numbers and the Central limit theorem of mathematical statistics, but it is different from them, so it cannot be directly applied to sampling technology. From a theoretical point of view, the importance of this part of the content is even stronger than the sampling organization forms. In order to help students realize the necessary knowledge transfer from mathematical statistics theory to sampling theory, it is necessary to supplement the basic contents such as "sampling distribution". In terms of teaching, these basic and important contents need to be carefully designed and carefully taught by teachers. Therefore, the proportion of these contents in classroom teaching should be increased, which is the "additive" realization of course teaching.

2) Do APP exercises after class, giving way to theory lectures and Q&A

In the Internet era, I tried to use class time to let students do exercises on the APP, which was a little unexpected - it took up more time. Only very important knowledge points will be practiced in

class, so every question needs to be explained by teachers in addition to students. Therefore, the author suggests that when using APP resources for teaching, all exercises on the APP should be kept after class, except for necessary classroom tests, which can be used in class for fairness, and the time in class should only be used for explanation. This can not only realize the classroom time "good steel to use the knife edge", but also facilitate the school to "mobile phone not into the classroom" implementation in the end, is conducive to improve the efficiency of students in class.

3) Online classes should not be too long

As a core course, in some universities it is the only major course in statistics that deals with surveys, and it is necessary to expand the content of the sample survey classroom appropriately. In the Internet age, it is logical that the necessary supplementary content should be taught online. For example, in addition to the complete teaching of classroom theory, practical contents such as questionnaire design and analysis, software operation content, extended cases and exercise explanation can be included in online teaching, but the duration should not be too long, which can only be an organic supplement to face-to-face teaching. Only offline and online linkage can make the sampling survey system perfect and students' listening, and self-learning realize mutual promotion.

(2) The computer operation needs to be carefully handled by the teacher

Today's statistics course teaching and related work can not be separated from machine processing. In the appendix of the classic textbook "Sampling Survey" (Fourth edition, by Jin Yongjin), the sampling organization forms of a complete case, the R language code from sampling to estimation and the necessary code explanation are clearly given. Even so, when it comes to getting students to do "textbook to script" code portability as an assignment, students still encounter a lot of problems. In the 2021-2022 school year of sampling survey teaching, the author conducted a convenient survey for students. Of the 31 students who participated in the survey, 70 percent of them said they hoped teachers could guide them on the computer. This fully shows that when using R language for sample extraction and estimation, even if students have software foundation and ready-made code, it is easy for students to "know what it is but not know why", so that new questions will not "change the code". The proportion of computers and what teachers teach are also an important aspect of optimizing the curriculum structure.

(3) Combining theory and practice using real data

The realistic and rich data available in R makes it easy for users to use real data for analysis exercises. The author programmed the sampling process and estimation of various sampling organization forms by using the data set of a certain R, which is a good supplement to the after-class exercises provided by classical textbooks.

(4) Bringing research-based learning into the classroom

In sampling theory lectures, the teacher will tell the students that repeated sampling can make the sample size greater than the total unit quantity, while no repeated sampling can not. Usually this sentence is passed around as a feature of repeated sampling, the teacher does not explain too much, and the students do not explore the specific use of this feature. With the support of R, teachers can make students' grasp of theory rise to the level of combining theory with practice by running the code, so that students can achieve "knowing and using":

```
# Randomly generate 10 genders (M, F):
Sample (C ("M", "F"), size = 10) # Do not repeat sampling, the result is an error
Sample (C ("M", "F"), size = 10, replace=T) # Repeat sampling to get the desired result
```

The premise of improving students' autonomous learning ability is to enhance students' interest in learning, and the teaching method combining theory with practice can greatly stimulate students' interest in learning, so as to realize the organic mutual promotion of "interest-exploration".

(5) Introducing flexible assignments to improve assessment methods

Assessment for learning involves fairness, equity, comparability and science, and needs to maximize impact on an achievable basis. The sampling class consists of online learning, offline learning, homework, software operation, and final paper. It is suggested that the final score be divided into three parts: the paper accounts for about 60%, and the topic is mainly the basic content of sampling theory and its application. Online learning and interaction takes up around 15%, with the remaining 25% being the operational content, which can be achieved through student videos with explanations. To avoid uniformity in the practical work, students can be divided into ten groups according to the last digit of their school number, with each group setting an estimated topic. Assignments that differ from each other are an important enabler of students' ability to improve their independent learning. There are many flexible homework departures, e.g., applying students' sequence numbers to generate random numbers to ensure that the data is not identical, etc., which will not be repeated here.

5 Conclusion

The introduction of software application into the course of "Sampling Survey" is the consensus of frontline teachers. The open source software R language, with its characteristics of statistical processing codification, greatly reduces the threshold for mastering it, so it is the best choice for the course of "Sampling Survey". This paper takes improving students' autonomous learning ability as the starting point, introduces the significance of R language for "Sampling Survey" from the aspects of data provision and visualization, and explores the reform of the existing teaching mode in detail. It can be believed that the combination of R and sampling theory is not only necessary for the teaching method combining theory and practice, but also necessary to enhance students' interest in learning, improve their hands-on ability and ability to solve practical problems. It will help improve students' autonomous learning ability and is worthy of promotion in the future.

Funding

This work was financially supported by grants from the ideological and political demonstration course for graduate students in Hebei province (YKCSZ2021142), research topic of party construction in colleges and universities of Hebei province (GXDJ2021B), research and practice of higher education teaching reform in Hebei province (2021GJJG631), educational reform project of Hebei GEO university (2019J09).

References

[1] Li Hao. 3D-oriented Reform and Practice of "Sampling Survey" Using Statistical Behavior Ability as the Core [J]. *Journal of Zhejiang Wanli College,* 2013.26(03): 105-108.

[2] Zhang Qinghong. Exploring the teaching reform of sample survey course [J]. *Modern Business Industry,* 2014.26(07): 128-129.

[3] Li Chunping. Reform and discussion on the practical teaching of "sample survey" course[J]. *China Electricity*

Education, 2014(09): 129-131.

[4] Hou Zhenmei. An innovative study on the experimental teaching of Sampling Survey [J]. *Modern Business Industry*, 2016.37(10):168-169.

[5] Huang LiFang, Wang Fang, Xu Lixia. The Application of Project Based Group Study to Improve Sampling Survey Course [J]. *Education and Teaching Research*, 2015. 29(12):79-83.

[6] Zhang Xuexin. Practice of Using Statistical Software Method to Teach Sampling Survey Curriculum [J]. *Journal of Ningxia Normal College*, 2015. 36(06):83-91.

[7] Zhu Song. SPSS Software is Applied in the Course of "Sample Survey" [J]. *Educational Teaching Forum*, 2019(22): 185-186

[8] Lu Yugui. The application of case teaching method in teaching the course of "Sample Survey"[J]. *Asia-Pacific Education*, 2015(04):104+93.

[9] Li Xiongying. A preliminary investigation on the application of statistics teaching based on R language[J]. *Journal of Higher Education*, 2017.

[10] Chen ZG, Wang Y. Teaching design of statistics based on R language[J]. *Tomorrow*, 2017.

[11] Lv Xiaokang. Exploration of psychological statistics teaching model based on R and RStudio [R], 2017, The 10th China R Conference (Beijing).

[12] Liu Yan, Li HongZheng. Reform and practice of Financial Time Series Analysis based on R language[J]. *Times Finance*, 2019.

[13] Zhao Peixin, Yang Yiping. R software implementation of random sampling without put-back in the teaching of "Sampling Survey"[J]. *Science and Technology Information*, 2012(31):26+28

[14] Wang Wei, Chen Zhijun, Xu Chenwu. A random sampling method based on R language and its application[J]. *Journal of Yangzhou University (Agriculture and Life Science Edition)*, 2014. 35(02):77-81.

[15] Lu Yugui. Teaching method of whole group sampling cases based on R software[J]. *University Education*, 2018(07): 65-67

[16] Qu Deqiang, Li Chencai. Analysis of the current situation of contemporary college students' data literacy and research on cultivation methods--a case study of a science and technology university in Shanghai[J]. *Research on Ideological and Political Education*, 2015. 31(04):97-100.

[17] Cao Fu. History of foreign education [J]. 1979, Beijing: *People's Education Press*

[18] Liu Jiansheng, Zhang Ping. The contemporary needs and construction priorities of accounting majors in the context of new liberal arts[J]. *Business Accounting*, 2021(19):120-122.