Curriculum Reform and Practical Application under the Background of Information Technology and Big Data—Take the Course of "Multivariate Statistical Analysis" as an Example

Manli Zhang*, Zhenhuan Kang, Zhimin Zhang

Taishan University, Tai'an City, Shandong Province, 271000, China.

Abstract

“Multivariate Statistical Analysis” is one of the compulsory courses for undergraduates majoring in statistics, big data and computer science. It is a branch of mathematical statistics that is most closely related to practical activities such as social economy, finance, psychology, medicine, sports and other fields, have a very wide range of applications. The content in the most used and popular multivariate statistics textbooks in colleges and universities today is relatively outdated. Many multivariate statistical methods generated under the background of big data are very practical, and they are not difficult to understand and operate, and can be provided to undergraduates stage of college students to study. Therefore, it is very necessary to add the method of multivariate statistics that is in line with the current data analysis company in the actual teaching, and it is also to reduce the degree of derailment of the ability of the actual company when college students graduate to work. Many methods of multivariate statistics under the background of big data can no longer meet the processing of modern data. Therefore, the latest research methods of multivariate statistics should be reflected in the actual teaching, so that students will not embark on the road of data analysis after graduation. Derail. In the actual process of teaching, it is planned to use a combination of flipped classroom and traditional classroom. In terms of teaching methods, by readjusting the time inside and outside the classroom, the knowledge point modules originally taught in the classroom are sent to the students in the form of video before learning the knowledge point, so that the students can use the time after class to learn independently. And with heuristic questions. Through out-of-class learning, students develop their doubtful knowledge points and proposed ideas in the actual teaching classroom together with the teachers. Therefore, this paper conducts research on the reform of teaching methods and teaching content in the course of multivariate statistical analysis.

Keywords

Teaching Content, Multivariate Statistical Analysis, Big Data

1. Introduction

The current era is the era of information explosion, which is accompanied by massive data. Many times we want to
obtain more and more in-depth information from data. The methods in multivariate statistics are actually very effective for data processing. The content in the most used and popular multivariate statistics textbooks is relatively outdated [1]. Many multivariate statistical methods generated under the background of big data are very practical, and they are not difficult to understand and operate. They can be provided to college students at the undergraduate level to learn. Therefore, it is very necessary to add the method of multivariate statistics that is in line with the current data analysis company in the actual teaching, and it is also to reduce the degree of derailment of the ability of the actual company when college students graduate to work [2]. Therefore, it is very necessary to realize the reform of teaching content in multivariate statistics teaching.

Secondly, in the actual teaching, most of the teaching methods are indoctrination methods, which cannot mobilize the enthusiasm of students well, and the learning efficiency of students is also low. And the students in the experimental class is only a simple realization of the textbook example [3]. In reality, the company is actually a form of teamwork, so the teaching method should change this situation, improve students' subjective initiative, and give play to students' teamwork ability. It is also imperative to reform the learning methods for the above problems.

2. Teaching method and content reform

In the actual process of teaching, it is planned to use a combination of flipped classroom and traditional classroom. Flipped classroom [1] is a new type of teaching mode. By readjusting the time inside and outside the classroom, the knowledge point modules originally taught in the classroom are sent to the students in the form of video before learning the knowledge point, so that the students can they use the time out of class to learn independently, with heuristic questions [4]. Through out-of-class learning, students develop their questionable knowledge points and proposed ideas in the actual teaching classroom together with the teachers. This can not only improve the enthusiasm of the students but also use the questions as a guide to enhance the subjective initiative of the students. Some knowledge points are really difficult, and need to be deduced with blackboard writing and face-to-face teaching. For such knowledge point modules, the traditional teaching mode is still used to ensure the quality of students' learning [5].

The main direction of reform in the teaching content is that many models in the textbook are flawed at the level of processing big data. On the premise of ensuring the learning of teaching content and teaching goals, with the data closer to the company's processing as a starting point, to exercise students' ability to analyze data. And by collecting data, selecting appropriate models, and interpreting model results to form different teaching cases. For example, principal component analysis is a very important chapter in multivariate statistical analysis. The multivariate statistics textbooks used by mainstream colleges and universities mentioned above do not introduce the application of principal components in image compression and recognition [6]. Clustering and image recognition are widely used, and these will be entered into teaching cases. Of course, the collected cases should not only meet the requirements, but also have practical significance and practicality. The establishment of the case practice teaching manual can enable students to experience the principles, ideas and processes of multivariate statistical analysis from the things and people around them, and realize the practicability of the knowledge they have learned, thereby stimulating students' interest in learning. Add new methods in the context of big data to the teaching content to shorten the degree of derailment between students and the company. The specific modules for adding content are shown in Figure 1.

Figure 1. Introduction of new methods and cases in the context of big data.
3. Practical application examples of multivariate statistical analysis

Adding practical examples in life to the case of multivariate statistics. Here are two examples. This is an application that is commonly used by companies but not in textbooks, and is a very practical application of multivariate statistical analysis in life. We expand on the example of multivariate statistics in image clustering and image compression.

First of all, let’s look at the application of principal components in image clustering and image compression. In real life, it is often encountered that the uploaded pictures are too large. The principal components can not only reduce the size of the pictures, but also take care of the clarity of the pictures. For specific operations, you can choose any picture. We choose a lakeside scenery as an example, and use R language to write the language based on the function `prcomp()`. The number of principal components is mainly selected as 3, 10, and 100 for comparison, and the specific diagram is shown in Figure 2. The third image is very close to the original image but its size is only 0.16 times the original image.

Next, let's take a look at the application of K-means clustering method in image color. The specific process selects any picture. Here we take a beautiful autumn picture as an example, use R language to program, and perform K-means clustering on the color. In the clustering process, K is equal to 2, 4.16, and 32, respectively. Let's check the results of the image: as shown in Figure 3; we can also get the palette in the clustering diagram as shown in Figure 4. In actual processing, you can also get the colors you want according to actual needs, and we will not show them one by one here.

![Figure 2. The lakeside image corresponds to the compressed image and the original image with principal components of 3, 10, and 100 respectively (arranged from top to bottom right row by row).](image-url)
4. Conclusion

The teaching mode and course content construction of the multivariate statistical analysis course have a great influence on the development trend of the related majors with data processing as the core application. Therefore, the content of multivariate statistics must be in line with the method of multivariate statistics in the company's data processing in reality. In the actual teaching, it must be close to the teaching of the latest methods, so as to better improve the students' data analysis ability in multivariate statistics.

Funded subject:

Taishan College Teaching Reform and Research Project (Project number: 4101210116).
References


