Research on the Application of Probability Statistics in Science and Engineering Courses Under the Background of Big Data

Yuanyuan Mao(✉)

College of Information Engineering, Shandong Management University, Jinan, Shandong, China
14438120191332@sdmu.edu.cn

Abstract. Probability and statistics is a highly applied subject, and its application in science and engineering courses can allow students to better understand and apply the theory and knowledge of professional courses. In order to better adapt to the training goals of compound talents, based on spss, matlab, python and other software, and based on the data in probability statistics, multiple models such as multiple regression models, nonlinear programming and ordinary differential equations are established. The results of the models are obtained respectively, and the results are analyzed in combination with specific cases to build a platform for students' data modeling, data simulation, data analysis and data mining. It can effectively cultivate innovative new engineering and new science talents with multi-disciplinary and broad vision, and provide talents and scientific and technological support for the country's economic and social development.

Keywords: Probability and Statistics · Course Integration · Matlab Software · Data Statistics and Analysis · Nonlinear Programming

1 Introduction

With the continuous progress of science, more and more probability and statistics theories and methods intersect with other disciplines, such as information science, mechanical engineering, biological engineering, physics, genetics, finance and so on [1, 2]. Probability and statistics, as an important university basic course, is the basic course of many emerging important subjects, and students of science and technology have systematically studied it before learning specialized courses. However, under the background of new engineering and science education, it is an inevitable requirement of higher education in the new era to cultivate compound knowledge talents [3]. The further integrated application of probability statistics in science and engineering courses is an important measure to transform talent cultivation from quantity to quality and improve students' innovative thinking and application ability.

With the advent of the era of big data, people are producing and accumulating massive data in various activities. The characteristics of the era of big data require new engineering and science talents not only to master solid professional basic knowledge

© The Author(s) 2023
C. F. Peng et al. (Eds.): EIMT 2022, AHSSEH 3, pp. 451–459, 2023.
https://doi.org/10.2991/978-94-6463-012-1_50

[4], but also to have the ability to apply probability and statistics theories and methods to solve complex problems in their professional fields [5]. Therefore, probability statistics occupies an important position in the new engineering and science curriculum system, and it is of great significance to study the application of probability statistics in science and engineering courses under the background of big data.

2 Application Problems

2.1 The Disciplines are not Closely Related

Probability statistics is closely connected with many science and engineering disciplines, but in the actual teaching process, the training programs and teaching programs of professional courses do not fully integrate probability statistics knowledge, and the existing programs and programs lack pertinence and practicability to cultivate students' comprehensive quality [6, 7]. Although probability statistics is a basic course for science and technology majors, it is difficult for students to master its connotation and essence, and it is difficult to integrate it with professional knowledge organically in the actual learning process.

2.2 Teachers of Specialized Courses Lack the Application Ability of Probability Statistics

Some teachers unilaterally believe that professional courses only need to teach professional knowledge. In the teaching of professional courses, they cannot deeply analyze the probability and statistics knowledge required by professional courses, and lack the necessary understanding of the application of probability and statistics. Probability and statistics teachers generally belong to the School of Science or school of Mathematics, and science and engineering teachers have few opportunities to communicate with them, and do little research on the teaching content and methods of probability and statistics knowledge integration, resulting in disconnection in the practical application of probability and statistics [8].

2.3 Lack of Probability and Statistics Experiments

The knowledge of specialized courses has a large amount of information and involves a wide range of subjects. In the case of tight class hours, teachers of specialized courses often do not carry out the experiment of probability statistics application in actual teaching. The experimental course is an important way for students to master the principles and logic of multidisciplinary integration, and an effective means to improve students' ability to find, analyze and solve problems. Through experimental operation, it is not only a consolidation and sublimation of the knowledge learned, but also a most direct way to cultivate students' application ability.

3 The Application of Probability Approach in Science and Engineering Courses

In view of the above application status of probability statistics in science and technology courses, it is necessary to improve both teaching and learning [9, 10]. One is how to teach professional knowledge while enabling science and engineering majors to master more practical probability and statistics methods. The other is how to improve students' self-learning ability and enhance their thinking and practice ability by using big data and Internet technology outside the classroom. In view of the first aspect, we should carry out a comprehensive reform from the teaching content of professional courses, and the selected probability and statistics content should be integrated with the professional knowledge of science and technology, so that the random mathematical method can be reflected in the professional knowledge. In view of the second aspect, we should reform the teaching methods and teaching ideas. On the premise of taking students as the center, we should adopt the mixed teaching method, through case teaching, experimental teaching, group discussion, simulation teaching, etc., to improve the interest in learning, stimulate the enthusiasm for learning, and train the students' application and innovation ability.

3.1 Course Content Suitable for the Cultivation of Compound Talents

The curriculum construction team was established to formulate the curriculum syllabus, emphasize the indispensable knowledge background of probability and statistics, and strengthen the supporting role of professional courses. The teaching syllabus is the action guide of personnel training, which relates to the achievement of personnel training goals and professional knowledge mastery standards. For different science and engineering majors, the specialty characteristics should be fully considered when they are combined with probability and statistics. In the original basis of inheritance in the development of new engineering, a new science professional, most of them have the characteristics of the professional knowledge is more abstract, logical, such as physical institute, school of mechanical engineering, control science and engineering institute, college of electrical engineering, college of information science and engineering and other professional, in the teaching content can be integrated into the knowledge of relevant data analysis ability. For the new engineering and science majors produced in the way of crossover and fusion, they have the characteristics of complexity, novelty and big data, such as the school of finance, business school, school of International trade, school of software and other majors, and the application of probability statistics should focus on data analysis and processing (Fig. 1).

3.2 Adopt the Online and Offline Teaching Mode

Under the background of big data, more and more colleges and universities begin to use the hybrid teaching mode combining online and offline. The blended teaching model changes the traditional teaching and increases the process of online learning for students. Through flipped classroom, MOOCs, mobile learning and other learning modes, students

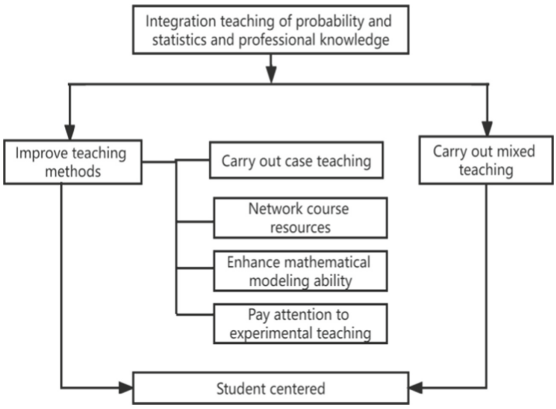

Fig. 1. Teaching reform of probability and statistics application in science and engineering courses.

can access and learn a large number of online open platform courses according to their own interests. According to their own different actual learning needs, students use a variety of knowledge learning mode to learn the theoretical knowledge of the course, and at the same time in the process of learning to develop interest in the knowledge, so as to further improve the learning effect. Through online learning, students can know in advance how well they have mastered the knowledge points they are about to study. For teachers, through the big data analysis of online learning platform, students' learning situation can be clearly known. Meanwhile, students can further watch relevant excellent videos to solve the puzzles of students and the knowledge points they do not understand about the application of probability and statistics. Finally, through these platform data, teachers can further carry out targeted teaching in the offline teaching process.

3.3 Enrich and Innovate Various Teaching Methods

Science and engineering courses have the characteristics of comprehensiveness and strong practice. When teaching basic principles and laws, teachers should vividly introduce specific probability and statistics cases in the discipline, and closely combine basic theory and probability and statistics practice to effectively improve teaching. quality. The problem-introduction teaching method is adopted to cultivate students' ability to discover problems, analyze problems, solve problems, and summarize and reflect. Facing the content that is highly abstract and difficult for students to understand in the teaching process, teachers should put forward corresponding questions in combination with the specific content. Using the group discussion teaching method, the class students are divided into several groups, and group discussions are carried out, so that students can deepen their understanding of what they have learned in the process of discussion. This teaching method can greatly improve students' enthusiasm for learning and activate the classroom atmosphere. Applying simulation technology to the teaching of science and engineering majors and building a virtual simulation experiment platform can better integrate with the application of probability and statistics knowledge.

3.4 Increase the Probability and Statistics Application Experiment Course

Experimental class is an important way for students to use professional knowledge and cultivate innovative consciousness. In the course of science and technology teaching, the use of computer platform and big data technology, increase the probability statistics application experiment course. The experiment content can be representative and practical experiment in the course, or it can be routine experiment. In order to adapt to the advent of the era of big data, colleges of science and technology should also establish laboratories combining professional knowledge with probability and statistics. Based on mathematical software such as SPSS and MATLAB, it builds a platform for students' data modeling, data simulation, data analysis and data mining. e.g.:

Regression model. Nutritionists study the effect of protein content in food on infant growth. The specific data are shown in Tables 1 and 2. Please fit the regression line of protein content in food on infant height and perform a significance test.

For this problem, it is a univariate regression problem, record height as y, age as x, establish model $y = \beta_0 + \beta_1 x + \varepsilon$, and estimate the parameters of the model through matlab software based on the data to obtain a specific model (Tables 3 and 4).

The confidence intervals of β_1 do not contain zero points, $p < \alpha$, R^2 is larger, indicating that the model has high accuracy, and according to the calculation results

Table 1. High-protein food groups

Age	0.2	0.5	0.8	1.0	1.0	1.4	1.8
Height/cm	54	55	63	66	69	73	82
Age	2.0	2.0	2.5	2.5	2.7	3.0	
Height/cm	83	80	91	93	94	94	

Table 2. Low-protein food groups

Age	0.2	0.4	0.7	1.0	1.0	1.3	1.5
Height/cm	51	52	55	61	64	65	66
Age	1.8	2.0	2.0	2.4	2.8	3.0	
Height/cm	69	68	69	72	76	77	

Table 3. High protein solution results

Regression coefficients	Regression coefficient estimates	Regression Coefficient Confidence Intervals
β_0	50.5224	[47.4739, 53.5709]
β_1	15.8976	[14.2577, 17.5376]
$R^2 = 0.9764$ $F = 455.2296$ $p < 0.0001$ $s^2 = 5.3810$		

Table 4. Low protein solution results

Regression coefficients	Regression coefficient estimates	Regression Coefficient Confidence Intervals
β_0	50.7934	[48.2292, 53.3575]
β_1	9.1884	[7.7351, 10.6417]
$R^2 = 0.9462$ $F = 193.6459$ $p < 0.0001$ $s^2 = 4.0949$		

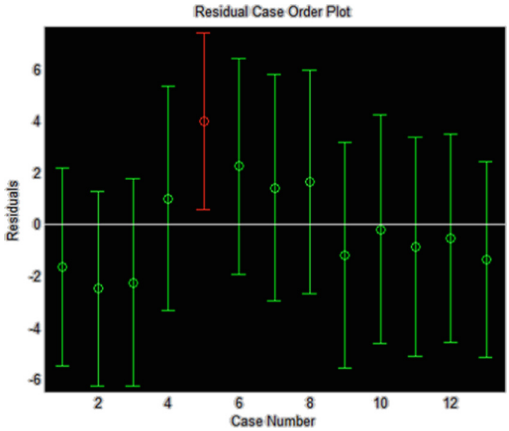

Fig. 2. Residuals and their confidence intervals for the high-protein food group.

(15.8976 and 9.1884), using the property test of $\hat{\beta}_1$, it has a significant impact. The regression lines of the high and low protein food groups are (Figs. 2 and 3):

$$y = 50.5224 + 15.8976\,x\,y = 50.7934 + 9.1884x \tag{1}$$

Non-linear planning: There are two existing rectangular pieces of land that need to be fenced to enclose the two pieces of land respectively. The area of each piece of land should not be less than 1000, and the height of the fence should not be lower than 2. Every brick that can be used to build a fence is the same, each brick has a height of 10, a length of 30, and a width of 15 (assuming the width of the brick is the width of the wall). The developer hopes to use 100,000 bricks to maximize the sum of the two circled pieces of land. How to enclose the land?

This problem belongs to the optimization problem. First, analyze the problem to determine the decision variables, take the length, width and height of the two rectangular areas as the decision variables, set the length of the first piece of land as a, the width as b, and the height as h_1; The length of the piece of land is c, the width is d, and the height is h_2; the maximum sum of the areas of the two circled pieces of land is s as the objective function. The constraints meet the various requirements in the title. The following nonlinear programming model is established:

$$\max \quad s = ab + cd$$

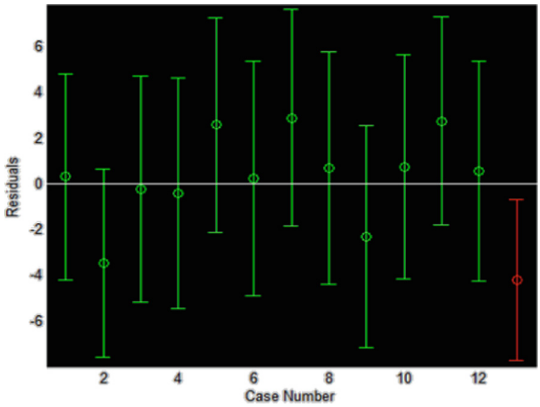

Fig. 3. Residuals and their confidence intervals for the low-protein food group.

$$s.t.\begin{cases} ab \geq 1000, \\ cd \geq 1000, \\ h_1 \geq 2, \\ h_2 \geq 2, \\ \frac{h_1}{0.1} \times \frac{2(a+b)}{0.3} + \frac{h_2}{0.1} \times \frac{2(c+d)}{0.3} \leq 100000, \\ a, b, c, d, h_1, h_2 \geq 0. \end{cases} \tag{2}$$

Using lingo software to solve, $a = b = 31.62$, $c = d = 343.38$, $h_1 = h_2 = 2$. Both pieces of land are square walls with a height of 2, and the side lengths are 31.62 and 343.38 respectively.

Data statistics and analysis. In the problem of street lamp replacement, considering that there is still a certain recycling value (constant) for the bulbs that are not damaged, establish the corresponding mathematical model and obtain the expression of the replacement cycle. When the average lifespan of a certain brand of bulbs is known to be 4000 h, subject to N (4000, 1002), the installation price of each bulb is 70 yuan, and the management department sets a penalty fee of 0.02 yuan/h for each unlit bulb. The recycling price of an undamaged light bulb is 5 yuan, and the optimal replacement cycle is calculated.

The life of the bulb $X \sim N$ ($\mu, \sigma 2$), the probability density function $p(x)$, the probability distribution function is $F(x)$, then $F(x) = \int_{-\infty}^{x} p(x)dx$.

Let the installation price of each light bulb be a, the penalty fee specified by the management department for each unit time (h) of each unlit light bulb is b, and the recycling price of each undamaged light bulb is c. Denote the replacement cycle as T and the total number of bulbs as K. The cost of installing the bulb is Ka, Penalty is $Kb \int_{-\infty}^{T} (T - x)p(x)dx$, $Kc \int_{-\infty}^{T} p(x)dx$ from recycling of undamaged light bulbs. Therefore, the average cost per unit time of the objective function

$$g(T) = \frac{\text{Ka} + \text{Kb} \int_{-\infty}^{T} (T - x)p(x)dx \text{ - Kc} \int_{-\infty}^{T} p(x)dx}{\text{T}} \tag{3}$$

Find T such that g(T) is the smallest. Let $\frac{dg}{dT} = 0$,

$$b\int_{-\infty}^{T} xp(x)dxTp(T) - a + c - c\int_{-\infty}^{T} p(x)dx = 0 \tag{4}$$

$$\int_{-\infty}^{T} xp(x)dx = \mu F(T) - \sigma^2 p(T) \tag{5}$$

$$\int_{-\infty}^{T} p(x)dx = F(T) \tag{6}$$

and get

$$(b\mu - c)F(T) + (cT - b\sigma^2)p(T) = a - c \tag{7}$$

According to the meaning of the question, $\mu = 4000$, $\sigma = 100$, a = 70, b = 0.02, $c = 5$, and the optimal replacement cycle obtained by python software is 4.1175e + 003. It can be seen that the original question remains unchanged in other conditions. In a certain range (c from 1 to 20), with the increase of the recycling price c, the optimal replacement cycle T becomes smaller.

In the era of big data, college students can acquire information quickly and in large quantities. Students expect that the knowledge they have learned can be put to practical use, and that they will be able to use the knowledge they have learned and software tools to solve practical problems., the integration of real life, and truly achieve the teaching purpose of applying what you have learned.

4 Conclusions

The application of probability statistics in science and technology courses is a topic worth exploring, especially in universities whose goal is to build first-class applied universities. In the era of big data, how to overcome the traditional teaching mode and use innovative teaching methods is both an opportunity and a challenge for teachers and students. Through the mixed teaching mode, case teaching, problem introduction teaching, group discussion teaching and simulation experiment teaching, the probability approach knowledge is deeply integrated into the science and engineering courses, which can effectively improve students' thinking ability and enhance the engineering practice ability and innovation ability required by new engineering and science talents. To cultivate high quality compound talents leading the development of future technology and industry.

Acknowledgements. This research is sponsored by the teaching reform project of Shandong Management University "Application Research of Probability and Statistics in Science and Engineering Courses in the Background of Big Data", the project number is YJG2020-79.

References

1. Huang, Chao. 2021. Probability and statistics course students application ability training research and practice. *Science Wenhui (ten-day)* 5: 65–66.
2. Liu, Kunlun. 2019. Exploration on the reform of probability theory and mathematical statistics course under the training mode of applied talents. *Journal of Qilu Normal University* 34(01): 28–33.
3. Li, Chunjuan. 2011. Teaching research and practice of college mathematics courses adapted to application-oriented talent training mode. *Journal of Lishui University* 33(05): 86–89.
4. Qi, Wuxia. 2016. Research on the cultivation of students' mathematical application ability in probability and statistics teaching of colleges and universities. *Statistics and Management* 7: 181–182.
5. Qi, Hongsheng, and Xichao Sun. 2018. Under the background of the new engineering applied probability and statistics in colleges and universities teaching exploration. *Science Tribune (late)* (18): 118–119.
6. Ren, Yuanyuan. 2021. Internet and big data in the application of probability and statistics teaching. *Integrated Circuit Applications* 38 (9): 196–197.
7. Wang, Fen, and Zhang Shaoyan, Xiaoying Wu. 2020. Reform and practice of probability theory and mathematical statistics courses under hybrid teaching mode. *Higher Education Journal* 28: 86–88.
8. Wang, Huili. 2019. Application oriented probability statistics teaching reform and innovation practice exploration. *Education Modernization* 26: 43–44 + 51.
9. Zhao, Xuefen. 2020. Research on teaching reform of probability theory and mathematical statistics guided by big data analysis ability. *Science Tribune (under the ten-day)* 30: 144–145.
10. Zhang Chunyun, Cui Chaoran, and Peiguang Lin. 2022. The theory of probability and statistics and its application teaching reform for the training of new engineering talents. *Computer Education* 02: 114–117 + 123.

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.