

Analysis of the Influence Factors on Undergraduate Specialty Choosing

Lulu Zhang

Faculty of Science
Ningbo University
Ningbo, China
1737739399@qq.com

Abstract—In all universities carrying out the policy of classified enrolment, professions distribution problems attend. To help the school provide reasonable and professional guidance about professional triage, we take university freshmen as the research object, and analyze the relationship between different categories of students and professional choice factors by correspondence analysis. Then provide scientific theory basis for the school, and combined with the specific circumstances of different students, provide more scientific and operable suggestions. The analysis results show that the eight categories can be divided into four types, each type of students have significant differences on different factors.

Keywords- large class enrollment; major selection; Correspondence analysis

I. INTRODUCTION

The beginning of 21 centuries, many universities recruit students by subject categories. Large class enrollment can reduce the blindness when students fill out college applications, and be conducive to cultivate comprehensive talents, so it has good prospects for development. Therefore, college students face the choice of professional issues and the schools face the problem about providing scientific guidance. So, there is a realistic meaning to study professional division problems.

Students will be affected by many factors in the independent choice of profession, including professional itself, the students themselves, family and schoolmates and social environment etc. In order to study the main influence factors which the different categories of students choose professions, this paper adopts correspondence analysis. Because influence factors are too many, we use the idea of dimension reduction and select the major influence factors. The schools can adjust the emphasis according to the analysis results when providing professional guidance. According to the different categories of students' needs, increasing the relevant guidance and helping students understand and choose professions better.

Correspondence analysis is a statistical technique to reveal the association between variables through interaction summary consisting of qualitative variables, and synthesize the advantages of R-analysis and Q-analysis. It's convenient to explain and infer something by graphs. Therefore, this paper choose graphical presentation, Transforming the contingency table into corresponding analysis diagram through correspondence analysis, and then it's visually to describe the

correlations between the different categories of students with different professional choice factors.

II. DATA ACQUISITION AND PREPROCESSING

This paper uses the method of questionnaire survey to obtain data, and the questionnaire includes four aspects on the students themselves, family and schoolmates, profession itself and social environment, totaling 21 contents. The concrete problems follow:

TABLE I FACTORS OF SELECTING THE PROFESSION

profession itself	students themselves	family and schoolmates	social environment
1.whether it is key;2.employment and wages;3.the difficulty of curriculum and academic requirements;4.the level and strength of teachers;5.the situation of going abroad or intercollegiate exchange;6.the ratio of going to graduate school;7.The current atmosphere of learning the profession	1.it fit my discipline specialty;2.it fit my hobbies;3.my grade is high, and I feel too wasteful if don't select it;4.I can select it according to present grade;5.reading it is more appropriate because I am a female (male) ;6.reading it is more appropriate because my constitution is better (worse)	1.my families work on this field, so I want to read it;2.family is wealthy and reading it is easier to go abroad;3.my family is poor, and the professional work is stable;4.parents' or relatives' proposal;5. friends' or classmates' advice	1. it is hot (cold) in modern society;2. the job is stable;3. the professional status is higher

The investigation object is a student who faces the independent profession choice and has stayed for a semester in school. The participation is voluntary online (all students are able to see the questionnaire), totaling 972 students completing the questionnaire. The measurement uses "Likert five-scale", including five levels, and they are "not very important", "not important", "general", "important", and "very important".

In the collected questionnaire, engineering students account for 24.2%; legal, literature and historical students account for 12.8%; economic management's students account for 21.2%; biochemical students account for 13.9%; natural science's students account for 9.1%; international business management's students account for 9.6%; construction's students account for 2.8%; and medical students account for

6.3%. The distribution of proportion coincides with the school's recruiting proportion.

III. CORRESPONDENCE ANALYSIS

A. The Principle of Correspondence Analysis

Correspondence analysis[1] is a multivariate statistical method which develops on the basis of R-analysis and Q-analysis. Correspondence analysis synthesizes their advantages and overcomes their disadvantages.

Correspondence analysis reveals the connections between the variables through interactive summary which consists of qualitative variables, the corresponding relationships between the different variables' various categories and differences between the same variables' various categories. But the key is that it can have the loads of variables and samples reflected factor on the same axis in the common, and link the variables with samples, and it's convenient for researchers to interpret and infer. It can get the internal relationships between several groups of data which can't see any contact seemingly.

B. The Corresponding Mathematical Model Analysis

The starting is that getting a probabilistic contingency table whose data comes from normalized original data. The table is a two-dimensional table consisting of rows and columns. Columns represent different things and rows represent the attributes of different things. According to the statistics data, we can get a contingency table as shown in table 2, and the cross point is the corresponding frequency.

TABLE II TWO VARIABLES IN CONTINGENCY TABLE

	B1	B2	...	Bp	Total
A₁	<i>x₁₁</i>	<i>x₁₂</i>	...	<i>x_{1p}</i>	<i>x_{1.}</i>
A₂	<i>x₂₁</i>	<i>x₂₂</i>	...	<i>x_{2p}</i>	<i>x_{2.}</i>
⋮	⋮	⋮		⋮	⋮
A_n	<i>x_{n1}</i>	<i>x_{n2}</i>	...	<i>x_{np}</i>	<i>x_{n.}</i>
Total	<i>x_{.1}</i>	<i>x_{.2}</i>	...	<i>x_{.p}</i>	<i>x</i>

Table 2 shows that x_{ij} ($i=1,2,\dots,j=1,2, n,\dots, p$) belongs to the A_i and B_j for the same time. To obtain the normalized probabilistic contingency tables, we can make $p_{ij}=x_{ij}/x$, then get a corresponding contingency table.

Step 1: We can get a matrix X of n rows and p columns from x_{ij} in table 2, and then make a corresponding transformation of the data. Later we get a matrix $Z=|z_{ij}|$ by calculating.

Thereinto

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

$$z_{ij} = \frac{x_{ij} - x_i \cdot x_j / x}{\sqrt{x_i \cdot x_j}} \quad (i = 1, 2, \dots, n; j = 1, 2, \dots, p)$$

Step 2: R-analysis: calculating the characteristic values ($\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$) of covariance matrix $A=Z'Z$ according to the cumulative percentage:

$$\sum_{\alpha=1}^m \lambda_{\alpha} / \sum_{\alpha=1}^p \lambda_{\alpha} \geq 85\%$$

Taking the m ahead eigenvalues ($\lambda_1, \lambda_2, \dots, \lambda_m$), and calculating the corresponding unit eigenvectors (u_1, u_2, \dots, u_m). Later we get a matrix F .

$$F = \begin{bmatrix} F_1 & F_2 & \cdots & F_m \\ u_{11} \sqrt{\lambda_1} & u_{12} \sqrt{\lambda_2} & \cdots & u_{1m} \sqrt{\lambda_m} \\ u_{21} \sqrt{\lambda_1} & u_{22} \sqrt{\lambda_2} & \cdots & u_{2m} \sqrt{\lambda_m} \\ \vdots & \vdots & & \vdots \\ u_{p1} \sqrt{\lambda_1} & u_{p2} \sqrt{\lambda_2} & \cdots & u_{pm} \sqrt{\lambda_m} \end{bmatrix}$$

Step 3: Q-analysis: calculating the unit eigenvectors $Z_{u1}=V_1, Z_{u2}=V_2, \dots, Z_{um}=V_m$ of matrix $B=Z'Z$ according to the m eigenvalues from step 2. Later we get a matrix G .

$$G = \begin{bmatrix} G_1 & G_2 & \cdots & G_m \\ v_{11} \sqrt{\lambda_1} & v_{12} \sqrt{\lambda_2} & \cdots & v_{1m} \sqrt{\lambda_m} \\ v_{21} \sqrt{\lambda_1} & v_{22} \sqrt{\lambda_2} & \cdots & v_{2m} \sqrt{\lambda_m} \\ \vdots & \vdots & & \vdots \\ v_{n1} \sqrt{\lambda_1} & v_{n2} \sqrt{\lambda_2} & \cdots & v_{nm} \sqrt{\lambda_m} \end{bmatrix}$$

Step 4: drawing a plane projection of factors: projecting the values of F_1, F_2, \dots, F_m and G_1, G_2, \dots, G_m on the same graph, then solving practical problems according to the calculations and distances between points and points.

C. Analysis of Factors Affecting College Students' Average Number Corresponding to the Professional Choice

This paper analyses the relationships between various factors by correspondence analysis. Firstly, taking the mean of different factors' influence, and then getting a two-dimensional contingency table[2]. Later, import the data into the SPSS[3] and doing corresponding treatment, then we get the result as shown in figure 1.

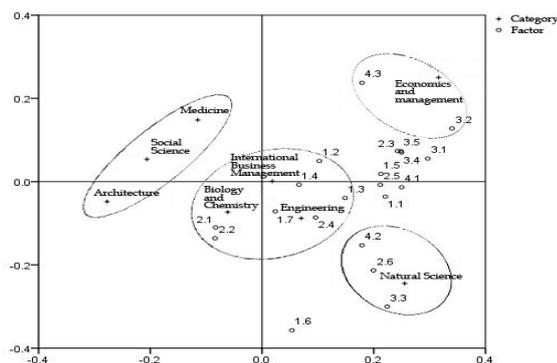


FIGURE 1: THE CORRESPONDING ANALYSIS DIAGRAM

Figure 1 shows that different categories of students present four clusters, each cluster highlights different characteristics. Medical students form a separate cluster. They take their physique, family and social economic situation into account, and tend to choose a stable job with a certain social status. International business management class students also form a separate cluster. They tend to choose a job which is easy to go abroad and has a certain social status. The students of biochemistry, economic management and engineering form a cluster, and the students are restricted when they choose the specialty because of grades. And they consider the difficulty of major courses and the teaching levels in addition to the employment situation and salary level. Individual interests and professional learning atmosphere are also important. The students of natural science, law, politics, literature, history and architecture form a cluster. The cluster doesn't include any kind of factors. It shows that the effect of each factor is smaller when the students choose professions.

IV. THE RESULT ANALYSIS

According to the results[4] of figure 1 and combined the results of the questionnaire with the actual situation, we make relative analysis for different categories of students.

From the questionnaire, we know that more medical students are from countryside, and in the cases of poverty and employment pressure, being a doctor is stable. And the medical profession has certain requirements on students' physical conditions (such as not achromatopsia). The Students of international business management are more affluent. They have some overseas plans and the occupational social status is relatively higher. From the analytical results, we know that medical students in the choice of profession is affected by the professional stability and body condition factors significantly, and the students of international business management think family conditions are conducive to go abroad and the occupational social status is higher. The statistical analysis fits actual situation, so the analytical conclusion is reliable through correspondence analysis. Therefore, in the study of the influence factors of students' professional choices, correspondence analysis is a reliable and meaningful analysis method.

The cluster of biochemistry, economic management and engineering is in the middle position of the coordinate,

indicating that students have certain tradeoffs on various aspects when choose profession. In fact, the number of the students in this cluster is maximal. The professional requirement is higher, and the influencing factors vary with each individual. The results correspond to the actual situation.

The rest of students are not affected by individual factors significantly. The data from the questionnaire can be seen that the differences are larger. Analyzing the actual situation, the students' professional purpose isn't clear, or they aren't satisfied with the profession.

V. ANALYSIS OF PROFESSIONAL GUIDANCE COUNTERMEASURE

Combined with the correspondence analysis map, the school can provide professional guidance according to the different categories of students' needs.

The students of medicine and international business management have clear goals. The school just provides further guidance on their career planning so that they can have more clear and more specific plans.

The students of biochemistry, economic management and engineering are affected by development prospects, employment situation and their own interests. Because the number of students in this cluster is more than half of the total number, and they consider roundly when choose professions. The school can provide different types of guidance for students to choose. On the one hand, the school can introduce the basic professional situation and the employment prospects according to the needs of students. On the other hand, the school can invite outstanding students to tell their own experience.

According to the results of correspondence analysis, we can know the rest of students are not affected by individual factors significantly. And the professional choices of the two categories of students are narrow, optional professions is not too many, and students have blindness in the choice of profession. The school can strengthen the education of career planning, and let them have more clear goals to choose their own professions.

ACKNOWLEDGMENT

My deepest gratitude goes first and foremost to my teacher, Weihua Li, for his constant encouragement and guidance. He has walked me through all the stages of the writing of this paper. Without his consistent and illuminating instruction, this paper could not have reached its present form.

REFERENCES

- [1] Xiaoqun He, multivariate statistical analysis, 3rd ed., Beijing: China Renmin University Press, 2011.
- [2] Weiguo Cheng, Wenxi Lu, and Lei Gong, "Application of correspondence analysis method in water quality evaluation of stone door in the reservoir", Water saving irrigation, 2011.
- [3] Wentong Zhang, Statistical analysis of SPSS advanced tutorial, Beijing: Higher education Press, 2004.
- [4] Feicheng Ma, and Dongmin Li, "Comparative study on Evaluation of difference analysis corresponding", Information Science, 2005.