THE STATISTICAL ANALYSIS OF POSTGRADUATE ENTRANCE EXAMINATION DATA BASED ON THE SAS SOFTWARE

E.I. Gubin Ph.D., Associate Professor
Chzhan Lifan, student gr. 8PM0I
Tomsk Polytechnic University
E-mail: lifan1@tpu.ru

Introduction

With the development of the times, there are more and more employment options for contemporary college students, including a series of options such as postgraduate entrance examination, job search, civil service examination and going abroad. Among them, the postgraduate entrance examination is undoubtedly a very popular option, either because they want to receive a higher level of education, or to avoid employment, or if they want to further improve themselves, a considerable number of people will choose the postgraduate entrance examination, and it will be derived from it. A series of peripheral industries have emerged. Postgraduate entrance examination counseling, postgraduate entrance accommodation and various postgraduate entrance examination institutions emerge one after another. The postgraduate entrance examination is undoubtedly a very popular topic in modern society. More and more college students are beginning to devote themselves to the torrent of postgraduate entrance examinations. Different people have different opinions on the postgraduate entrance examination. In an environment where the surrounding students choose the postgraduate entrance examination, if they do not have a clear plan and a clear understanding of their future, they will be easily affected by the choices of others and may not A decision that suits you. In short, the postgraduate entrance examination is a topic that not all college students can avoid. Therefore, in such a large environment, it is a very valuable and meaningful thing to analyze the results of the postgraduate entrance examination.

Method Description

Logistic regression, also called Logistic regression analysis [1], is a generalized linear predictive regression model, which is often used in data mining, disease diagnosis, and economic prediction. Logistic regression is the most popular dichotomous data model, and its general form is as follows:
The ordinary Logistic regression model is:

$$p_i = \frac{\exp(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\cdots+\beta_m x_{im})}{1+\exp(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\cdots+\beta_m x_{im})}, i = 1,2,\cdots,n. \quad (1.1)$$

There are only two categorical values for the dependent variable y in the logistic regression model, usually 0 and 1. Suppose that in the independent variables x1, x2,..,xn under the action, the probability of an event occurring is p, and the probability of not occurring is 1-p, $\frac{p}{1-p}$ is the ratio of the probability of occurrence to the probability of non-occurrence, recorded as the superior ratio (odds). If we take the natural logarithm of odds, we get:

$$Logit(p) = \ln(odds) = \ln\frac{p}{1-p}, \quad (1.2)$$

Called the Logit transformation of p, the transformed Logistic regression model is:

$$Logit(p) = \ln\frac{p}{1-p} = \beta_0 + \beta_1 X_1 + \cdots + \beta_i X_i + \varepsilon. \quad (1.3)$$

Table 1. Part of the discipline entrance exam national line in 2016

Subject category	Total score	Single subject (full score = 100 points)	Single subject (full score = 150 points)
Science	285	39	59
Engineering	265	36	54
Managent	335	45	68
Economis	325	45	68
Literature	350	53	80
Jurisprudce	315	45	68

Logistic Regression /Analysis Result

According to the results of parameter estimation, whether the six variables of political score, professional course 2 or high math score, professional course 1 score, type of school applied for, whether to take maths, and whether to apply for this school are consistent with this event There is strong statistical significance [3].

The logistic regression model expression is:

$$p = \frac{\exp(-12.8756+0.0602x_1+0.06x_3+0.0408x_4-1.5887x_6+1.5646x_8+1.5186x_{10})}{1+\exp(-12.8756+0.0602x_1+0.06x_3+0.0408x_4-1.5887x_6+1.5646x_8+1.5186x_{10})} \quad (1.4)$$

It can be seen from the expression that the coefficients of the variables x1, x3, and x4 are very small, and basically do not have much influence on the dependent variable. The analysis of the odds estimation result is shown in Tab.2.

Table 2. Odds ratio estimations

Optimization estimate			
Effect	Point estimate	95% Wald confidence limit	
x1	1.052	1.028	1.097
x3	1.062	1.041	1.083
x4	1.042	1.024	1.060
x6	0.204	0.081	0.516
x8	4.781	1.824	12.532
x10	4.566	1.395	14.946

The odds ratio represents the ratio of the probability of an event occurring to the probability of an event not occurring, which can be taken in all non-negative regions. If the odds value is 1.5, it means that the possibility of the event is unlikely to happen 5 times. The estimated odds ratio represents the effect of the independent variable on the probability of an event. The estimated odds ratio is greater than 1, which means that the probability of the event will increase, or in other words, the independent variable has a positive effect on the probability of the event. The estimated odds ratio is less than 1, which means that the probability of the event will decrease, or in other words, the independent variable has a negative effect on the probability of the event. The odds ratio estimated value is 1, which means that the independent variable has no effect on the probability of the event.

Conclusion

Through SAS software, analyse the results of the postgraduate entrance examination, and through Logistic regression and cluster analysis, we can get the level of influence of each factor on the success rate of a voluntary postgraduate entrance examination. It can provide some basis for college students when choosing a college to apply for, and do it when filling in their volunteers. Make a reasonable choice.

Reference

1. He Xiaoqun, Liu Wenqing. Application regression analysis［M］. The third edition. Beijing: Renmin University of China Press, 2011.
2. Wang Xuemin. Application of multivariate analysis［M］. The third edition. Shanghai: Shanghai University of Finance and Economics Press, 2009.
3. Wang Haibo, Luo Li. SAS statistical analysis from entry to master [M]. Beijing: People's Posts and Telecommunications Press, 2013.