Multivariate Statistical Analysis of Japanese Entrance Examination Data   II

Haruo YANAI[1] , Mituaki HUZII[2], Norio SUZUKI[3]
National Center For University Entrance Examination
1)&3) Research Division, 2)Vice President
2-13-28 Komaba Meguro, Tokyo, Japan
1)yanai@rd.dnc.ac.jp, 2)huzii@cen.dnc.ac.jp 3)suzuki@dnc.ac.jp

Shingo SHIRAHATA
Dept. Of Math. Science, Osaka University, 1-3, Machikaneyama Toyonaka, Osaka, Japan,sirahata@sigmath.es.osaka-u.ac.jp

1. University Admission System in Japan

In Japan, all of the national and regional/local universities and some private universities select their students on the basis of the scores of National Center Test (NCT) and the scores of examinations conducted by individual universities. The main purpose of NCT is to assess the level of applicant basic achievement abilities acquired in upper secondary school. Many universities require 5 subjects on NCT. Every year more than half million applicants wishing to enter university take NCT. Applicants taking 5 subjects are obliged to take Japanese Language(Jap), Mathematics , (Mat)and. Foreign Language mainly English ,and in addition they must choose one subject from Geography and Histories, and also one subject from Sciences. It is to be noted, here, that all questions are objective and are answered using a mark sheet.

2. Data analyzed in this study

The purpose of the presentation is to offer the results of analysis which intend to look into the multidimensional structures of the achievement abilities required to solve five main subjects of the NCT. For the purpose, we used the data of the NCT conducted in January,2000. In order to compute the correlation coefficients among any pair of five subjects, we classified the data into the following six groups; ie;

Group1: Jap,Mat,Eng,Wh and Ph(14,930)   Group2:Jap,Mat,Eng,Jh and Ph(17,248)
Group3: Jap,Mat,Eng ,Ge and Ph(57,476)   Group4:Jap,Mat,Eng,Wh and Bi(26,832)
Group5 :Jap,Mat,Eng ,Jh and Bi (33,717)   Group6:Jap,Mat,Eng,Ge and Bi(25,816)

Where Wh, Jh, Ge, Ph and Bi stand for World history, Japanese history, Geography, Physics and Biology, respectively. We showed the numbers of the data for each of the six groups in parenthesis.

3: Methods of Analyses and Main Results

3.1 Principal Component analysis of the correlation matrix of the five subjects

First, we conducted the principal component analysis(as abbreviated PCA hereafter)to six correlation matrices of the five subjects, using the data of all the cases mentioned above.

From the analysis, it was shown that the contribution ratios of the first principal components ranged from 59% to 66% and the principal loadings of the five subjects in all six groups were positive, thus establishing that the first PC measures the basic achievement abilities required in solving all items of the five subjects.

As for the second and third PC, it was rather difficult to find the common names of the principal components extracted from each of the six groups.

Next, we performed PC to the correlation matrix based on Jap, Mat, and Eng, which are all

compulsory to applicants taking five subjects on NCT. The contribution ratios of the first PC ranged from 67% to 72%, and thus it can easily be estimated that the first PC corresponded to exactly the basic dimension measuring the basic achievement abilities of the NCT.

3.2 Principal component analysis of the five subjects eliminating the effect of total achievement abilities

Thus, we tried to compute the partial correlation matrices of the five subjects by eliminating the effects of the first principal factor(f) based on the main three subjects. Mathematically, we compute the partial correlation matrix as follows,

$$PCOR(X'X/F) = D^{-1/2}(R_{XX} - r_{xf}(r_{xf})')D^{-1/2} \quad where \quad D = Diag(R_{XX} - r_{fx}(r_{fx})')$$

and Rxx is the correlation matrix of the five subjects and r_{fx} is the vector of correlations of the principal factor f and each of five subjects. The Analyses using the data of all the six groups produced the first PC whose contribution ratios were smaller than those from the PCA of the original variables 2, while on the other hand the contribution ratios of the second PC are bigger, since the partial correlation matrix resulted from eliminating the general achievement abilities required to solve all of the five subjects (See, Figs 1 and 2).

Let us interpret what each extracted PC implies. The PCA of the partial correlation matrix computed for each of the six groups confirmed that the first PC measured the dimension reflecting Mathematics versus Japanese, and the second PC measured dimension reflecting English versus the remaining four subjects. In this connection, it is interesting to note that Japanese language and English and mathematics are shown to measure complete different ,provided that the level of the general achievement abilities is fixed constant for each applicant taking NCT.

4: Discussions

In order to estimate the total achievement abilities, we computed in this study the principal component analysis of Japanese Language, Mathematics and English. Further, it can be possible to estimate the total achievement abilities by means of the sum score of the five subjects, or the sum score of only two subjects such as Mathematics and Japanese, and so on. We will show these results in the presentation, and consider what is the most appropriate method to estimate the total achievement abilities using the data of NCT.
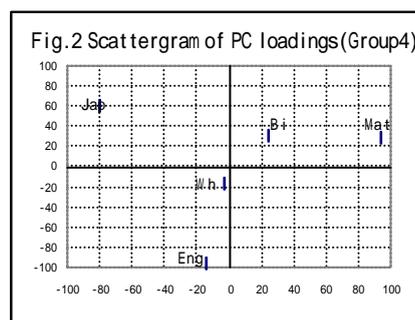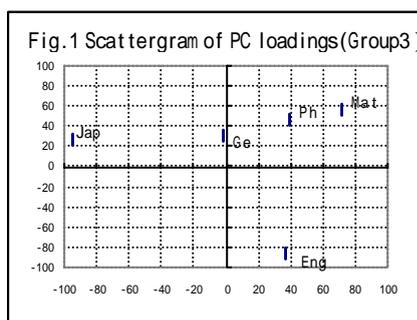

Fig.1 and Fig.2   Principal Component loadings of the first and second PC of the five subjects (based on Group3 and Group4