Development of Item Analysis and Related Programs for Personal Computer

SAEGUSA, Norio.

ABSTRACT

A series of personal computer programs were developed in order to carry out statistical analyses of the outcome of the examination/quiz/placement–test with the emphasis on assessing the quality of questions given in those examinations. Programs are designed to be executed on conventional 16–bit personal computer.

1. INTRODUCTION

In competitive test, only important information is the point of individual student and the order of the points irrespective to the quality of the questions. Therefore statistical analyses are made on points. Those analyses have little significance for the test which is made to assess the achievement of the student. Quality and appropriation of the questions should be carefully examined so as to assess the achievement accurately. Such analyses of the questions of the test may, in some extent, be made with item analysis[1,2].

In the field of Japanese language education for the students whose mother tongue, cultural background and the way of thinking are different from those of Japanese language teachers, sometimes a "good" question for the teachers appears to be "not good" one. In other words, the criteria of "good" questions are different for the students and for the teachers. The author's belief is that objective judgment deduced with certain statistical analysis may sometimes, even though it may not be always, provides a help for the teacher to develop a "better" test.

Present short note describes a series of personal computer programs developed to carry out statistical analyses on 16–bit personal computer, namely PC9801 series of NEC Corp.

2. Description of the Programs

Brief description of the function, the structure and remarks in executing the program, where appropriate, are given in this section.

2—1. Program Language

Program language used is N88BASIC(86)*1 which is operated on MS-DOS*2 operating system. Programs are initially written in N88BASIC interpreter, which provides easier debugging facility of the programs. After successful debugging, the N88BASIC interpreter source program is compiled to generate machine language code program, which ensures faster execution of the program, with BASICC*1 compiler program. BASIC language is available on almost all personal computers. Therefore the present interpreter source programs can be easily transferred to other personal computer system with minimum modifications.

Programs are executed under MS-DOS operating system. Therefore data exchange can be, in principle, made with other programs run under MS-DOS. Data files generated for the present programs may readily be used in other commercially available software, in some cases, after proper modification of the data structure.

2—2. Data Files

In order to analyze the test data with a personal computer, the results of the test should be available as the data file on the computer system. All the answers given by the students for the questions of the test and correct answers are referred as "data". In conventional competitive test result analysis, points of the students are considered as the basic data, i.e. treating the test result as a set of correct-or-wrong data irrespective to the nature of wrong answer One of the important objectives for the present analysis is that assessment of the quality of the test itself to establish better language education environment. In due course of such study, analysis of wrong answering would become important. Therefore the answers given by each student are recorded.

2—2—1 · Data File Name

Each data file should be identified with a "file name". Access of the data file is made by calling the name by the program. MS-DOS poses some restrictions on file name. Well classified file names should be selected for easier identification of the data file. Data files to be used in the present analyses are named according to the following rules;

i) first four characters should represent the identifier of the series of the tests, for example P86A, where P represents *p*lacement test, 86 for 19*86* and A for *a*utumn,

ii) following three characters should represent the type of the specific test subject, for example LIS represents for *lis*tening and so on.

File access by the below described data analysis programs is made by using those file names. Therefore above described rules for naming of the files should be strictly observed, otherwise the programs cannot identify the required data files for desired analysis. MS–DOS allows up to eight characters for file name and up to three characters for file name extension. File name extension is used to identify the type of the data file, which will be described below.

2 –2–2. Data Format

Data entry and some data manipulation, such as sorting, were made with dBASEIII [*3] in early stage of the program development. Therefore, the data format used in the present programs is, in certain extent, posed constraint for easier data manipulation with dBASEIII. In order to minimize error in the data entry, following data format was chosen. In order to distinguish four different selection of correct answer and no–answer, five different symbols should be used. Most straightforward candidate would be either five numbers or five alphabets. In both cases, one key action is required by selecting a key out of five different keys. On the other hand, four key actions are required with only two keys for "1000" format Main reason to use this answer–data–format is to minimize error in data entering process. The author believes that "1000" format is more time consuming but yields less error in data entering process. One of the reasons would be that data format conversion from actually used alphabetical– or numerical–answering format to "1000" format may always awake the person who enters the data The author believes that no error should be made in data entering process and every effort should be made to eliminate incorrect entering of the data. The "1000" format might require more patience of the person who enters the data, but overall time and tediousness, which might be encountered at the time of data correction, should be less with this data format.

Each student is identified with student code, which consists of four–digit number. First two digit represents the student's native country and the last two digit represents the serial number for the same country.

2 –2–3. Data Entry

Correct answers for the test are separately entered to form a correct–answer–file Correct answers for the test are entered with program called ANSINA, which generates the data file with ex-

tension of ANS. Answers given by each student are entered with dBASEIII. Data file generated with dBASEIII, which has DBF extension in the file name, is then transformed to an ASCII form data file with using "COPY" utility command of dBASEIII, since special data file structure is employed for the data file of dBASEIII. Those transformed data files have TXT extension in their file names.

It is most important to have correct data for any kind of data analysis. Most straightforward method to detect incorrect data entry is made by inspecting of the listing of the entered data file. This procedure was indeed employed in the very beginning of the analysis and appeared to be tiresome and time consuming. More practical error checking routine provided in the present analysis procedure is made by comparing the more than one data files which are generated by different persons. The program called COMPA was developed for this purpose. The program assumes that the entered answer is correct when all of the answers for the question entered in the different data files are the same On the contrary, any one of the answers in the data files is different from the rest of the answer (s), the program considers that the incorrect data entry was made and wait for correction. Usually three different data files are generated for this purpose. The author's assumption is that the probability to make incorrect data entry for the same question by the different persons should be negligibly small However it should be noted that "negligibly small" does not mean "no error". Therefore every effort should be exercised to enter correct data when one initially enter the data. It would be time consuming in the initial instance but total time for data analysis including error checking should be less.

Introduction of the mark sheet together with an on-line mark sheet reader would simplify the data entry, provided that the students having had good experience with mark sheet manipulation. If this is not the case, introduction of the mark sheet would put unnecessary strain on the students and should not be introduced. Such strain should be posed on the teachers when they enter the data.

2—3. Total Marks : MLIST

Total mark for each student can be determined with the program "MLIST" by comparing the answers of each student, which are recorded in TXT file, and correct answer data in ANS file. MLIST lists the marks of the students in the order of the student code and in descending order of total mark with student codes and names.

2—4. IF statement

Brief description of IF statement of BASIC is given here IF statement is used in COMPA and

MLIST to make conditional branching in the program, namely to judge whether the answer is correct or not by comparing the answer with the correct answer. This function is used in other programs to be described below. Statement has a format;

IF xxx = yyy THEN nnn (ELSE mmm)

where, if xxx = yyy is correct, execution jumps to the line number nnn, or to the line number mmm or jumps to next line, when xxx≠yyy, xxx and yyy are either numerical or character-parameter.

2-5. Item Analysis

So far two programs were developed to make item analysis. One of them was developed to calculate the parameters which are presented in reference (1) and was called ITEMA. Second one named ALPHAA and calculates the parameters described in reference (2). Detailed descriptions for statistical significance and derivation of the parameters are not given here and should be referred to the respective article.

2-5-1. DIS/DIF Item Analysis:ITEMA

Two parameters are deduced with ITEMA, i.e. DIS representing the order of distinguishability and DIF order of difficulty, and are defined as,

$$DIF = (NH+NL) / 2 \times N$$

$$DIS = (NH-NL) / 2$$

where NH and NL are the numbers of the students in the higher and lower grade group, respectively, giving correct answer for the question to analyze, and N is the number of the students in each group. N is determined with an empirical convention for statistical analysis. This statistical analysis is referred as DIS/DIF analysis hereafter.

Criterion for a good question according to DIS/DIF analysis is that all the students in the high-group give correct answer (NH = N) and none of the students in the low-group give correct answer (NL = 0), i.e. DIS = 1.0 and DIF = 0.5. Easy questions can be answered correctly by all the students in the high-group and part of the students in the low-group. Difficult questions are answered only by part of the students in the high-group. It is implicitly assumed that all the students in intermediate group give correct and wrong answer for easy and difficult question, respectively. Those

results are represented as;

easy questions being defined as NH = N and $0 \leqq NL < N$;

$$DIS = -2 \times DIF + 2 \text{ for } 0.5 \leqq DIF \text{ and } 1 > DIS \geqq 0$$

difficult questions as NL = 0 and $0 \leqq NH < N$;

$$DIS = 2 \times DIF \text{ for } 0 \leqq DIF < 0.5 \text{ and } 0 \leqq DIS < 1$$

However for more realistic situation, some of the students in the high-group may fail to give correct answer for an *easy* question. It is also probable that some of the *difficult* questions are correctly answered by the fraction of the students in the low-group. For those cases, DIS becomes smaller and DIF becomes either smaller or larger. It should be noted that DIS and DIF are calculated only with part of the data available. Therefore some of the information might be lost in DIS/DIF analysis, i.e. answering profile for the intermediate group is not taken into account and might affect the determination of "actual" distinguishability and difficulty of the question.

Data files for the high- and low-group are generated with conditional COPY utility command of dBASEIII. Student code of the students with higher or lower than respective reference points are collected to form new data files, which are identified as ffffHIGH. TXT and ffffLOW. TXT, where ffff is four-character examination identifier described in subsection 2-2-1.

ITEMA prints out NH, NL, DIS and DIF values for item analysis. Also it prints out the numbers of the students to give respective choice of each questions for high-, low-group and for all students, for "Good-Poor Analysis".

2-5-2. α Coefficient and Point-biserial Correlation Coefficient

Item analysis made with ITEMA deals only part of the data available and might loose some of significant information. ALPHAA deals with the data of all the students available. Parameters defined and described by Kaiho[2] are calculated with ALPHAA. "Pass Ratio" for the j-th question in the test, π_j, which might correspond to DIF in some extent, is defined as,

$$\pi_j = \frac{1}{N} \sum_{i=1}^{N} U_{ij}$$

where N is the number of the students, U_{ij} is either 0 or 1 for the i–th student making wrong or correct answer for the j–th question. Order of distinguishability is defined with point–biserial correlation coefficient, which is defined as,

$$r_j = \frac{\bar{X}_j - \bar{X}}{\sigma} \sqrt{\frac{\pi}{1-\pi_j}}$$

where $\bar{X}$ is the mean value of the points of the test, $\bar{X}_j$ the mean points for the students who make correct answer to the j–th question, σ standard deviation of the points for all students. Uniformity of the test, alpha coefficient α, which is used to determine whether the all questions in the test are designed to determine ability for the same element in whole spectrum of the subject, is also calculated. α coefficient is defined as,

$$\alpha = \frac{m}{m-1} \left\{ 1 - \frac{\sum_{j=1}^{m} \sigma_j^2}{\sigma^2} \right\}$$

where m is the number of the questions, σ_j^2 the variance of the points for the j–th question and σ^2 variance of the total points for the test.

ALPHAA prints out alpha coefficient, point–biserial correlation coefficient for each question and some of the conventionally referred statistical parameters, such as variance, standard deviation and so on. Output is self descriptive.

3. Concluding Remarks

Above described programs have been used to analyze the outcome of the placement test held at the Education Center for Foreign Students of the University of Tsukuba. The number of the samples for the placement test has been around 90 for previous three occasions. It should be noted that the number of the samples is not large enough for conventional sense of statistical analysis. Therefore strictly speaking obtained results are not considered as "statistical data". However obtained results can be utilized for assessing the quality and appropriation of the questions in some extent, even if it is not full extent, with careful data interpretation. Accuracy of assessment would become higher after accumulation of more data. Study and development of the method for objective assessment of quality and appropriation with small number of sample itself is important subject to be carried out Therefore assessment of different "statistical" parameters to be used for further analysis should be studied in the due course.

As mentioned earlier, some of the deta manipulation is carried out with dBASEIII. File conver-

sion process is necessary from dBASEIII format to ASCII format. This operation is rather time consuming and sometimes induces annoying mistakes, especially when generating high– and low–group files in DIS/DIF analysis. This function should be replaced with BASIC program in near future to avoid unnecessary confusion. Also initial data entry routine, which is now made with dBASEIII, should be developed. The author still believes dBASEIII being useful for the aid of program development and for testing the outcome of the developed programs. Any comments and suggestions for the existing and/or possible programs will be appreciated.

Acknowledgments

The programs described in the present article have been developed after close communication with some of the members of the Education Center for Foreign Students of the University of Tsukuba. The author thanks to Ms. Reiko SAEGUSA and Ms. Takako SAKAI for their comments on the programs and for test runs of the programs The author also thanks to Prof. Kazuo Otsubo for allowing him to access the personal computer software owned by the Center.

References

*[1]PC9801, N88BASIC (86) and BASICC are the trademark of NEC Corp.

*[2]MS–MOS is the trademark of Microsoft Inc.

*[3]dBASEIII is the trademark of Nihon Ashton Tate Inc.

(1) NIHONGO–NORYOKU–CHOSA–NINTEI–IIN–KAI ed., "GAIKOKUJIN–NO–TAMENO–NIHONGO–NORYOKU–NINTEI–SHIKEN–NI–KANSURU–CHOSA–KENKYU–NO–KEIKA–HOUKOKU V" (in Japanese), The Society for Teaching Japanese as a Foreign Language, 1985, Tokyo.

(2) H.KAIHO, "SHINRI–KYOIKU–DETA–NO–KAISEKIHOU–10–KO" (in Japanese), Fukumura Shuppan, 1985, Tokyo.