COUNTING RESPONSES TO MULTIPLE-CHOICE QUESTIONS
Charles A. DePascale and Margaret R. Rost
Advanced Systems in Measurement and Evaluation

INTRODUCTION

One of the first steps in analyzing results from any standard questionnaire or multiple-choice test is to compute the number and percentage of respondents selecting each response option to each question. The technique described in this paper shows how this task can be easily accomplished using a multi-dimensional array.

The code presented here will apply this technique to computing the number and percentage of respondents within a school district selecting each response option to a 20-item test with four options per item.

PREPARING THE DATA

The data file containing a district identifier and the individual student responses to the 20 items is read into a data set. The data are routinely scanned or key-entered so that the response options A, B, C, and D are coded 1, 2, 3, and 4, respectively. Often, however, the data file will contain missing values for items which the student did not attempt to answer and other out of range values to indicate that the student selected more than one response option. These invalid values must be cleaned before the data are processed in an array.

Using the following statements the response array is edited so that all invalid values are assigned a value of 5. In this way the number and percent of missing/invalid responses can also be computed.

```
ARRAY RESP{20};
DO I = 1 TO 20;
  IF RESP{I} IN (1,2,3,4) THEN RESP{I} = 5;
END;
```

CREATING AND INITIALIZING THE ARRAY

A two-dimensional array to count the number of students selecting each option is created - the first dimension representing the number of items (20) and the second dimension representing the number of response options (5). For each district, the array elements are initialized to starting values of 0.

```
ARRAY NRESP{20,5};
IF FIRST.DISTRICT THEN DO;
  DO I=1 TO 20;
    DO J=1 TO 5;
      NRESP{I,J} = 0;
    END;
  END;
END:
```

COUNTING THE RESPONSES

The responses to the 20 items are then counted using the following loop:

```
DO I = 1 TO 20;
  NRESP{I,RESP{I}} = NRESP{I,RESP{I}} + 1;
END;
```

The value of RESP{I}, 1 to 5, is resolved and becomes the second dimension of the array NRESP. Thus, the appropriate element of the array is incremented by 1. Consider, for example, the case where a student selected option B to question 1. The value of I would be 1 and the value of RESP{I} would be 2. The element NRESP{1,2} therefore would be incremented.

Because the values of the array NRESP are retained, all but the last record within the district can be deleted using the statement

```
IF LAST.DISTRICT EQ 0 THEN DELETE;
```

After all of the student records have been processed, the data file consists of a single record per district containing the number of respondents selecting each of the 100 possible response options.

COMPUTING PERCENTAGES

After the number of respondents selecting each option has been determined, the percentage of respondents can be computed easily. First the total number of students within the district is computed by summing the number of respondents to any item.

```
```
TOTAL = SUM(OF NRESP1-NRESP5);

An additional array is created and the percentages are calculated in the following loops:

DO I = 1 TO 20;
  DO J = 1 TO 5;
    PERC[I,J]=ROUND((NRESP[I,J]/TOTAL)*100);
  END;
END;

Each record in the final data set, therefore, will now contain the district identifier and the number (NRESP1-NRESP100) and percentage (PERC1-PERC100) of respondents selecting each of the 5 response options to each of the 20 questions.

CONCLUSIONS/LIMITATIONS

The technique described in this paper provides a quick method for computing simple questionnaire or test item results. One potential limitation of the technique is that the use of two-dimensional arrays makes more demands on memory. However, working on a PC we have been able to analyze the results of up to 15 items with approximately 50,000 respondents. For a 30 item test, we simply ran the program twice and merged the results.

This technique is best-suited to simple, straightforward analyses where frequencies of every response option are desired. It is not as useful for more complex analyses such as those in which a) response options are combined within an item, b) response options are combined across items (e.g. people who selected Option A to question 1 and Option C to question 3), or c) missing/invalid values are not included in the computation of percentages. To handle those situations we have developed other techniques making use of output data sets from multiple PROC FREQ procedures.