Behavior Research Methods & Instrumentation
1975, Vol. 7 (5), 475-476

An interactive computer program to optimize data categorization

GIDEON VIGDERHOUS and PAUL OELTJEN
Survey Research Group, Bell Canada, Montreal, P. Q., Canada

Description. This program was written in order to enable the researcher to categorize univariate distribution by minimizing the amount of information lost resulting from such a procedure. This program will indicate which categories should be collapsed by minimizing the information lost, given that the number of categories for collapsing the data are supplied by the researcher.

The mathematical discussion and procedure for data categorization is discussed (Bryson & Phillips, 1975) and is achieved by dividing a frequency distribution f(x) into L homogenous groups where boundaries b_1, $b_2 \circ \circ \bullet b_{1-1}$ are placed in such a way that:

$$\sum_{A}^{b_1}\sqrt{f(x)} = \sum_{x=b_1+1}^{b_2}\sqrt{f(x)} = \bullet \bullet \bullet = \sum_{x=b_{L-1}+1}^{B}\sqrt{f(x)},$$

where A and B denote the lowest and highest values of x.

Program Language. Fortran IV.

Computer. The program was originally written and tested for Honeywell 6000.

Restrictions. Since the initial input data is categorized, no problems are anticipated in information input. This program is written in an interactive mode; however, with few modifications, the program could be adjusted to other computer systems.

Program Input. (1) Number of categories of the original data set. (2) The observed frequency for each category. (3) The desired number of categories for reclassifying the data.

Program Output. (1) The observed frequency for each category x. [f(x)]. (2) The square root of the frequency for each category $\sqrt{f(x)}$. (3) The cumulative distribution of $\sqrt{f(x)}$. (4) The optimal division of categories (old categories should be

```
*RUN INTERVAL

PROGRAM TO OPTIMIZE DATA CATEGORIZATION
---------------------------------------

TO START A NEW RUN WITH NEW DATA
TELL THE PROGRAM TO COLLAPSE YOUR DATA INTO  0 CATEGORIES

TO TERMINATE THE RUN
TELL THE PROGRAM TO COLLAPSE YOUR DATA INTO -1 CATEGORIES


NUMBER OF CATEGORIES OF ORIGINAL DATA?
=10

ENTER THE OBSERVED FREQUENCY FOR EACH CATEGORY.
=12,23,34,5,56,12,3,14,10,23


OLD
CATE-      F        SQRT(F))    CUM(SQRT(F))
GORY
  1       12.       3.46410        3.46410
  2       23.       4.79583        8.25993
  3       34.       5.83095       14.09089
  4        5.       2.23607       16.32695
  5       56.       7.48331       23.81027
  6       12.       3.46410       27.27437
  7        3.       1.73205       29.00642
  8       14.       3.74166       32.74808
  9       10.       3.16228       35.91036
 10       23.       4.79583       40.70619


INTO HOW MANY CATEGORIES DO YOU WISH TO COLLAPSE YOUR DATA?
=7

  NEW          OLD
CATEGORY    CATEGORY    FREQUENCY

   1          1 -  1        12.
   2          2 -  3        57.
   3          4 -  4         5.
   4          5 -  5        56.
   5          6 -  7        15.
   6          8 -  9        24.
   7         10 - 10        23.


INTO HOW MANY CATEGORIES DO YOU WISH TO COLLAPSE YOUR DATA?
=4

  NEW          OLD
CATEGORY    CATEGORY    FREQUENCY

   1          1 -  2        35.
   2          3 -  5        95.
   3          6 -  7        15.
   4          8 - 10        47.


INTO HOW MANY CATEGORIES DO YOU WISH TO COLLAPSE YOUR DATA?
=-1

READY
```

```
*LIST INTERVAL

10 DIMENSION F(200),S(200)
20 WRITE (6,601)
30  601 FORMAT(//"PROGRAM TO OPTIMIZE DATA CATEGORIZATION"/
31&"---------------------------------------"//"TO START A NEW RUN WITH
 32& NEW DATA"/"TELL THE PROGRAM TO COLLAPSE YOUR DATA INTO  0 CATEGORIES
 34&"//"TO TERMINATE THE RUN"/"TELL THE PROGRAM TO COLLAPSE YOUR DATA
36& INTO -1 CATEGORIES")
40C
50C GET INPUT
60 1 WRITE (6,602)
70 602 FORMAT(//"NUMBER OF CATEGORIES OF ORIGINAL DATA?")
80 READ 50,N
90 50 FORMAT(V)
100 WRITE(6,603)
110 603 FORMAT(/"ENTER THE OBSERVED FREQUENCY FOR EACH CATEGORY.")
120 READ 50,(F(J),J=1,N)
130C
140C PRINT ORIGINAL DISTRIBUTION, SQRT(F), AND CUM (SQRT(F))
160 WRITE(6,604)
170 604 FORMAT(///"OLD"/"CATE-",4X,"F",8X,"SQRT(F))",3x,"CUM(SQRT(F))
175&"/"GORY")
180 SUM=0.
190 DO 10 I=1,N
200 SRF=SQRT(F(I))
210 SUM=SUM + SRF
220 WRITE(6,61)I,F(I),SRF,SUM
230 61 FORMAT(I4,F8.0,F14.5,F15.5)
240 10 S(I)=SUM
250C
260C START COLLAPSING RUN
270 11 WRITE(6,605)
280 605 FORMAT(///"INTO HOW MANY CATEGORIES DO YOU WISH TO COLLAPSE YOU
 285&R DATA?")
290 READ50,NEWK
300 IF(NEWK)99,1,2
310 2 SINT=SUM/NEWK
320 WRITE(6,606)
330 606 FORMAT(//3x,"NEW",8X,"OLD"/"CATEGORY",3X,"CATEGORY",3X,"FREQUE
 335&NCY"/" ")
340 KNEW=0
350 TEST=SINT
360 KLO=1
370C
380C DETERMINE NEW CATEGORY
390 21 DO 20 I=KLO,N
400 IF(S(I),GE,TEST)GO TO 22
410 20 CONTINUE
430 22 KHI=I
440 IF((S(I)-TEST),GT,(TEST-S(I-1)))KHI=I-1
450 IF(KHI.LT.KLO)KHI=KLO
460 31 FNEW=0.
465 KNEW=KNEW+1
470 DO 30 I=KLO,KHI
480 30 FNEW=FNEW + F(I)
490 WRITE(6,62)KNEW,KLO,KHI,FNEW
500 62 FORMAT(I5,I9," -",I3,F10.0)
510 IF(KHI.EQ.N)GO TO 11
520 KLO=KHI + 1
530 41 TEST=TEST + SINT
540 IF((TEST-S(KHI)),LT,(S(KLO)-TEST))
545&GO TO 41
550 IF(TEST.LT.SUM) GO TO 21
560 KHI=N
570 GO TO 31
580 99 STOP
590 END
```

collapsed into new categories). (5) The frequency observed for each new category after recollapsing the data.

Availability. The program printout and user instruction may be obtained without cost from:

G. Vigderhous
Bell Canada
620 Belmont, Room 1120
Montreal, P. Q., Canada.

REFERENCE

BRYSON, K. R., & PHILLIPS, D. P. Method for classifying interval-scale and ordinal-scale data. *Sociological Methodology*, 1975. Pp. 171-190. San Francisco: The Jossey Bass Series, 1974.

Behavior Research Methods & Instrumentation
1975, Vol. 7 (5), 476

CMTG: A Fortran IV Program for studying individual behavior in the common target game

FRANK A. MORISANO and VICTOR REZMOVIC
Syracuse University, Syracuse, New York 13210

The common target game was devised as a method for studying group problem solving (Leavitt, 1960). The game requires the members of a group to work together to solve problems under conditions of limited information. Each of three group members must contribute a number from 0 to 10 toward a target number between 3 and 30, previously specified by the experimenter. Most recently, Joyner (Joyner, 1970; Joyner & Green, 1970; Joyner, Note 1) has used computer techniques in developing and testing an information processing model of concept formation by individuals playing the common target game. Computerized procedures would appear to offer an effective way for studying the behavior of individuals interacting under highly controlled conditions. CMTG was written as a medium for conditioning subjects to solve problems presented according to specified solution strategies under either a cooperative or a competitive set (Morisano, Note 2).

Description. CMTG is an interactive computer version of the three-person common target game. The subject, working at a remote terminal, is presented with a description of the task and instructions for playing the game. He is informed that the responses of the other two members of his group will be determined by the program and that the goal of the group will be to develop a system for solving any target number on the first attempt. A series of randomly ordered target numbers is presented. The subject enters his contribution and then receives feedback concerning the responses of the two simulated members of his group, whether the group was successful in hitting the target on that trial, and the number of points earned. The responses of the simulated subjects are selected from among three general solution strategies referred to as TWOs, THREEs, and TENs (Joyner, 1970). Each of these strategies involves three different roles for partitioning target numbers. Each group member must adopt one of the roles if the group is to succeed. Points can be earned on either an individual or a group contingency basis. It is thus possible to require the subject to learn to respond according to up to nine different strategy-role combinations to be successful. The criterion for learning of a strategy-role combination is seven consecutive correct target number solutions. A new combination is automatically introduced when the subject reaches criterion for the previous one. In the present version, the game terminates following exposure to nine combinations.

Input. Before the subject is seated at the terminal, the experimenter must initialize a set of parameters that determine the strategy-role combinations the subject will receive, the order in which these will be presented, and the type of feedback provided (individual or group contingency).

Language. CMTG is written in Fortran IV and was developed on a DEC System-10.

Availability. Copies of the source program, a more elaborate description of it, and copies of sample output are available without charge from the authors at the Department of Psychology, Syracuse University, Syracuse, New York 13210 or from Frank A. Morisano, Board of Cooperative Educational Services, Box 233, Yorkville, New York 13495.

The authors would like to express their appreciation to Sidney J. Arenson and Edward J. O'Connell for their assistance in the development of CMTG.

REFERENCE NOTES

1. JOYNER, R. C. *COMTARG: Computer simulation of the common target game.* Toronto: FAS, York University, 1969.

2. MORISANO, F. A. Conditioning of individual behavior in the common target game. A dissertation proposal, Syracuse University, 1974.

REFERENCES

JOYNER, R. C. Computer simulation of individual concept learning in the three-person common target game. *Journal of Mathematical Psychology*, 1970, 7, 478-514.

JOYNER, R. C., & GREEN, C. J. Demonstration of computer-augmented group problem solving. *Behavioral Science*, 1970, **15**, 452-462.

LEAVITT, H. J. Task ordering and organizational development in the common target game. *Behavioral Science*, 1960, **5**, 233-239.