

A computer program for the generation  
of Hudson's Variance Index for  
categorical clustering performance

LeROY CLINTON and JUDITH ARTER  
University of Illinois, Urbana-Champaign  
Champaign, Illinois 61820

**Description.** Hudson and Dunn (1969) have suggested that categorical clustering performance data based on the Bousfield and Bousfield (1966) expected clustering index should be converted to standard score form, before it is statistically analyzed. Hudson and Dunn (1969) specifically advance this position for comparisons across differing numbers of semantic categories, or under conditions where independent or task variables differentially influence the recall of items. Use of this approach minimizes the confounding of treatment effects and procedural influences.

While Hudson and Dunn (1969) and Barton and David (1957) have presented a formula for the determination of the necessary variance values, the hand calculation of a variance value for each subject in each trial can become tedious. To facilitate the use of the Hudson approach, a computer program has been developed to generate a table of categorical clustering performance variance values. The program is based on the Hudson and Dunn (1969) and Barton and David (1957) formula (1) and yields variance values for strings of categorical responses.

$$\text{Variance} = \frac{F_2 (N-3)}{N(N-1)} + \frac{F_2^2}{N^2(N-1)} - \frac{3F_3}{N(N-1)},$$

where  $F_2 = \sum_{i=1}^k M_i (M_i - 1)$

$$F_3 = \sum_{k=1}^k (M_i - 1) (M_i - 2)$$

$M_i$  = number of words recalled in the  $i^{\text{th}}$  category.  
 $N$  = total number of words recalled.

The table requires up to 10 terms, Group 1 -- Group 10, to locate a variance value for a given observation of categorical clustering task performance. Group 1 is the number of recalled words in the verbal category with the most recalled words. Group 2 is the number of recalled words in the second largest recalled category. Group 3--Group 10 are similarly determined, where Group 10 is the number of recalled words in the least recalled verbal category. Obviously, if there are only six categories, Group 7--Group 10 will be zero in all cases. Likewise, if there are only four categories, Group 5--Group 10 will be zero in all cases.

Table 1

Number Recalled in Each Group										Total Recalled		
Group										(N)	E(R)	Variance
1	2	3	4	5	6	7	8	9	10			
6	5	5	3	3	3	3	3	3	1	35	3.029	2.656
6	5	5	3	3	3	3	3	3	2	36	3.000	2.648
6	5	5	3	3	3	3	3	3	3	37	3.027	2.690
6	5	5	4	0	0	0	0	0	0	20	4.100	3.164
6	5	5	4	1	0	0	0	0	0	21	3.905	3.020
6	5	5	4	1	1	0	0	0	0	22	3.727	2.891

For illustration, consider the case where there are six categories and the subject recalls six items in one category, five items in two more categories, and four, one, and zero items in the last three categories, respectively. In this case, ordering from highest to lowest, Group 1 = 6, Group 2 = 5, Group 3 = 5, Group 4 = 4, Group 5 = 1, Group 6 = 0. In addition, Group 7 -- Group 10 are also zero, since there were only six groups. The appropriate Hudson variance value for these parameters is shown in Table 1, which is a partial listing of the entire table. Thus when a subject recalls six, five, five, four, one, and zero items, the variance value for his performance is 3.020. In addition, the appropriate Bousfield and Bousfield (1966) expected clustering value (E(R)) is paired with each variance entry. The expected clustering values are based on (2).

$$E(R) = [(M_1^2 + M_2^2 + \dots + M_k^2)/N] - 1$$

Where  $M_i$  is the number of items recalled in a given category  $N$  is the total number of words recalled.

**Hardware.** The program is written in level-four Fortran, for an IBM 360 system.

**Input form.** In its present form, the program requires that jobs be submitted in card form.

**Availability.** Program listings available at no cost from the authors. However, this is limited to a small supply of copies.

Requests should be sent to LeRoy Clinton, Department of Special Education, 1005 West Nevada, University of Illinois, Urbana-Champaign, 61801.

REFERENCES

Barton, D. E., & David, F. N. Multiple runs. *Biometrika*, 1957, 44, 168-178.  
Bousfield, A. K., & Bousfield, W. A. Measurement of clustering and of sequential constancies in repeated free recall. *Psychological Reports*, 1966, 19, 935-942.  
Hudson, R. L., & Dunn, J. E. A major modification of the Bousfield (1966) measure of categorical clustering. *Behavior Research Methods & Instrumentation*, 1969, 1, 110-111.