Behavior Research Methods & Instrumentation
1976, Vol. 8 (4), 390

Analysis of variance with APL/360

STEPHEN MADIGAN
University of Southern California, Los Angeles, California 90007

This note describes a function written in APL/360 that computes an analysis of variance table for many kinds of experimental designs (excluding those with unequal n or missing observations). The only argument required by the function is the name of the array containing the data to be analyzed. No declaration of type of design whatsoever is needed. Instead, the Subjects factor is included as a dimension of the data array, an approach described by Lindman (1974, p. 189) as treating Error (subjects as a source of variability) "as a random factor nested in all factors in the design." Beyond this, the function also treats all data arrays as if they were fully crossed designs, with n = 1. In these respects, the function is similar to a FORTRAN routine written by Ogilvie (Note 1), and is also a generalization of computational methods described by Clifford (1968).

As an example of the use of this function, consider a classification with factors A (two levels), B (three levels), and Subjects (four). The data are assigned to X, a 2 by 3 by 4 array (see Figure 1). A call to the function (APLAOV X) produces a table of sums of squares, degrees of freedom, and mean squares for the following "effects:" A, B, AB, S, AS, BS, ABS. The function automatically assigns the letters A, B, C, . . . , to the first, second, third, . . . , dimension of the data array, with the letter "S" assigned to the last dimension. (The data array must be structured so that Subjects is the last dimension.) Denominators of F ratios would be formed from the last four terms by the user. If, for instance, the design was completely randomized (four subjects in each of the six cells), addition of the last four sums of squares, and division by the sum of their degrees of freedom, would produce the usual mean square within cells (assuming fixed effects for A and B). If factor A was between subjects and factor B within subjects, combining the S and AS terms would give the error for testing A, while the BS term alone would be the error for B, etc. In this latter case, the four levels of the Subjects factor would refer to the number of subjects in each level of A, or to the total number of subjects if the design was entirely repeated measures, with N = 4.

The function would also produce the necessary sums of squares if the design was hierarchical, if, for instance, factor B was groups nested under A, with four subjects in each nested group. In this case, the sums of squares for the nested factor would be gotten from A and AB.

The procedure generalizes to designs with any number of bases of classification, effectively limited only by workspace size. (Arrays up to rank 63 can be defined in APL.) At the cost of requiring the user to form error terms and perform the final F tests, the function achieves considerable generality and ease of use, due largely to the array definition and manipulation capabilities of APL. It has also proven quite useful didactically. Large ANOVA programs that produce not only F ratios, but p values as well, may not be entirely desirable in a first course analysis of variance, where the student might profit more from having to examine sources of variability in performing F tests.

```
    ∇ APLAOV X
[1]   V←((⍴S),NCL)⍴S←O←(1+NF←(2*(NCL←⍴LV←⍴X))-1)⍴0
[2]   VN←NCL⍴'ABCDEFG '
[3]   VN[NCL]←'S'
[4]   K←¯1
[5]  TK:D←X
[6]   CF←(BV=0)/⍳⍴BV←⌽BV←(NCL⍴2)⊤K←K+1
[7]   O[J]←+/1=V[J←K+1;]←BV
[8]   C←RD←0
[9]   →((⍴CN←VN[CF])=0)/SQ
[10] TC:D←+/[(VN⍳CN[C])-((C←C+1)-1)] D
[11]  →(C<⍴CF)/TC
[12] SQ:S[J]←(+/((×/(⍴D))⍴D)*2)÷(×/LV[CF])
[13]  RD←RD+S[1]×¯1*O[J]
[14]  →(K=0)/TK
[15]  N←1
[16] TN:→(N=(⍴LO←(O[⍳K]<O[J])/⍳⍴O[⍳K]))/SV
[17]  SN←¯1*(O[J]-O[LO[N←N+1]])
[18]  RD←RD+SN×S[LO[N]]×O[LO[N]]=(+/(V[LO[N];]=1)∧(V[J;]=1))
[19]  →TN
[20] SV:MS←(SS←S[J]+RD)÷DF←×/(LV[EF←(V[J;]=1)/⍳⍴V[J;]]-1)
[21]  VN[EF];'      ';SS;'      ';DF;'      ';MS
[22]  →(K<NF)/TK
    ∇
```

	b_1	b_2	b_3
a_1	3	5	4
	5	3	4
	3	2	6
	10	6	3
a_2	6	4	5
	10	11	10
	10	13	16
	3	13	15

APLAOV X

A	160.2	1	160.2
B	10.58	2	5.292
AB	31.08	2	15.54
S	58.83	3	19.61
AS	56.17	3	18.72
BS	22.42	6	3.736
ABS	76.58	6	12.76

Figure 1. APL analysis of variance function and example of use.

Language: APL/360 on a time sharing terminal. If APL-SV is available, the summary table may be more neatly formatted with little modification of the present version of the function.

Requirements/limitations. The function requires approximately 1,100 bytes. This allows analysis of fairly large classifications of data in the normal 32K-byte APL/360 workspace. The version of the function shown in Figure 1 will accept up to seven factors or bases of classifications, not including Subjects. This can be extended by adding to the list of letters (HIJ . . .) in the third line.

Availability. A listing, examples of use, and some description of the algorithm are available without cost from the author.

REFERENCE NOTE

1. Ogilvie, J. *How to use subroutine ANOVA.* Department of Psychology, University of Toronto, 1969.

REFERENCES

Clifford, T. Simplified computational programming for analysis-of-variance designs with correlated observations. *Psychological Bulletin*, 1968, **69**, 439-440.

Lindman, H. R. *Analysis of variance in complex experimental designs.* San Francisco: Freeman, 1974.