Analysis of variance with APL: Improvements and extensions

MARC HAMOVITCH

Navy Personnel Research and Development Center
San Diego, California 92152

The two programs described here represent modifications and extensions of an APL/360 analysis of variance program written by Madigan (1976). The first program, shown in Figure 1, while essentially producing the same information as Madigan's, represents improvements in several areas. Changes involving internal workings of the computer include reduced storage requirements, reduced processing time, and reduced overall execution time. As the size of the data array increases, the percent savings increases as well. Execution time for analysis of a 2 by 2 by 10 data array is reduced by 43%; for a 2 by 2 by 2 by 10 array, the running time is reduced 58%; for a 2 by 2 by 2 by 2 by 10 array, the time saved is over 80% or from over 30 min running time down to less than 6 min.

Further changes include a more easily readable summary table and a simple interactive means of calculating F values. The program calls for an interactive approach to provide for selection of the appropriate error terms. To allow for any type of design, the subjects variable is included as a random factor in the analysis nested within all of the other factors. As Madigan (1976) states, the data array is further treated as a fully crossed

E				
3	4	6	7	6
5	6	6	7	7
4	6	8	8	7
8	10	10	7	11
2	3	4	4	5
3	5	6	3	4
9	12	12	8	10
9	7	12	11	12

ANOVA E			
SOURCE	SS	DF	MS
1 A	.6250	1	.6250
2 B	192.2750	3	64.0917
3 AB	50.6750	3	16.8917
4 S	34.6500	4	8.6625
5 AS	3.7500	4	.9375
6 BS	14.3500	12	1.1958
7 ABS	26.4500	12	2.2042
ENTER SOURCE LINE #S OR 0 TO EXIT <u>2</u>			
ENTER LINE #S FOR ERROR TERM <u>4 5 6 7</u>			
F(3,32) = 25.896			
MSE = 2.475			
ENTER SOURCE LINE #S OR 0 TO EXIT <u>2</u>			
ENTER LINE #S FOR ERROR TERM <u>6 7</u>			
F(3,24) = 37.701			
MSE = 1.7			
ENTER SOURCE LINE #S OR 0 TO EXIT <u>1 2</u>			
ENTER LINE #S FOR ERROR TERM <u>4 5 6 7</u>			
F(4,32) = 19.485			
MSE = 2.475			
ENTER SOURCE LINE #S OR 0 TO EXIT <u>2</u>			
ENTER LINE #S FOR ERROR TERM <u>3</u>			
F(3,3) = 3.7943			
MSE = 16.892			
ENTER SOURCE LINE #S OR 0 TO EXIT <u>0</u>			

Figure 2. Equal-cell data array and example of ANOVA function. Four different F calculations for Factor B are displayed, showing the procedure for four different experimental designs. Underlined portions represent input.

```

V ANOVA X;CP;D;DF;DFF;DFS;I;J;L;LN;LV;MS;
MSE;NF;NS;NV;P;S;SP;SS;VN
[1] SOURCE SS DF MS
[2] SS+DF+S+(J+0)*NV+P+P*(NF+0)+0, (NS+
(2*(NF+P*LV+P*X))-1
[3] VN+((NF-1)*'ABCDEFB'), 'S'
[4] DUM:D+X
[5] 1+1+P*SP+(0=P*(J+1));J)/\NF
[6] +(J+NS+1)/SQ
[7] SUM:D+/(SP+1-1) D
[8] +(I>1)/SUM
[9] SQ:SQJ+((+/,D*2)+x/LV\SPJ
[10] +(J=1)/DUM
[11] MS+(SS\J)+S*(1*NV\J)-NV)*(+/P)=+/2=P
+(P)*P*P;J)=DF\J)+x/LV\CP+P*J;J)/\NF)-1
[12] 3+J-1;8+VNECP; 14 4 5 0 14 4 +SS\J,DF
J,MS
[13] +(J=NS)/DUM
[14] FT:(1=LN+1)+34+344LN+M,M+ENTER SOURCE
LINE #S OR 0 TO EXIT ')/0
[15] LD+1+1+29+29+LD+M,M+ENTER LINE #S FOR
ERROR TERM
[16] 'F(DFS;','DFE;') = '+(+SS\LN)-DFS+
DF\LN)-MSE+(+SS\LD)-DFE+DF\LDJ
[17] 'MSE = 'MSE
[18] -FT
V

```

Figure 1. Generalized equal-cell analysis of variance program.

The opinions and assertions contained herein are those of the writer and are not to be construed as official or reflecting the views of the Navy Department.

design with $n = 1$. This method has also been used in repeated-measures designs (Keppel, 1973).

Figure 2 shows an example of a data array (E) with Factor A (two levels), Factor B (four levels), and five subjects per cell. The factors are labeled by the function, with "S" always the last factor. The only restrictions for the data array are that subjects must be entered as the last dimension and every cell must contain data (unequal cell frequencies are treated by the second function to be discussed). The size of the data array is only limited by the work space available on the particular machine being used.

By typing the names of the function (ANOVA) and data array (E), the user begins the analysis. As shown in Figure 2, the function prints out the line number, source code letter, sums of squares, degrees of freedom, and mean squares. After the program has completed the analysis and printed the table, it solicits line numbers for the source term (numerator) then the error term (denominator) for the calculation of F. To illustrate how to use this part of the function, Factor B is analyzed according to four different designs. The first example shows how B is analyzed in a fixed-effects between-

subjects design. To indicate Factor B, "2" is entered as the source line number. The within-groups error term is the combination of all sources containing S as a factor. Thus, 4, 5, 6, and 7 are entered for the error term. The function then responds with the degrees of freedom, the F value, and the mean square error. In the next example, B is analyzed as a repeated-measures factor in a fixed-effects model. The proper error term is $B \times S/A$, or the combination of Lines 6 and 7. The third example shows how a nested design is handled. Factor B is nested within A with both factors fixed. To arrive at the B/A term, Lines 1 and 2 are combined. The final example involves a design in which Factor A is random and Factor B is fixed. The proper error term for B in this case is the A by B interaction represented by Line 3. To terminate the function, assuming all calculations have been made, "0" is entered as the source line number. In a similar fashion, any of the sums of squares can be combined to appropriately analyze any design except those with unequal n.

To fill the need for a program that would provide the correct analysis for data with unequal n, a second function (UWMAOV) was written (Figure 3). This algorithm was based on the unweighted means procedure described in Keppel (1973). It follows the same basic procedures as the ANOVA program and requires data to be organized in basically the same way. The program requires, however, that where zeros are used to represent real data, a constant must be added to all data points and zeros reentered to indicate missing data, so that zeros only represent the absence of data. Due to the complicated nature of the unweighted means analysis, the generality of the function is somewhat limited. The

U				
3	4	6	7	0
5	6	6	7	7
4	6	8	8	0
8	10	10	7	11

2	3	4	0	0
3	5	6	3	0
9	12	12	8	0
9	7	12	11	0

SOURCE	UWMAOV U	SS	DF	MS
A		.0618	1	.0618
B		160.9651	3	53.6550
AB		44.6458	3	14.8819
ABS		70.8500	25	2.8340
THE F VALUES ARE: WITH DF:				
A		.022	(1,25)	
B		18.933	(3,25)	
AB		5.251	(3,25)	

SOURCE	UWMAOV E	SS	DF	MS
A		.6250	1	.6250
B		192.2750	3	64.0917
AB		50.6750	3	16.8917
ABS		79.2000	32	2.4750
THE F VALUES ARE: WITH DF:				
A		.253	(1,32)	
B		25.896	(3,32)	
AB		6.825	(3,32)	

Figure 4. Unequal-cell data array and two examples of UWMAOV function. Underlined portions represent input.

function produces only one error term, the within-groups mean square. This term is used to automatically generate F values for all source terms. These values are then printed out with their associated degrees of freedom. Figure 4 shows Data Array U with missing values indicated by zeros and the analysis produced by the function. An analysis of Data Array E is also presented to show the similarity between the two functions. The only differences in the table printed by UWMAOV are that the within-groups error term is the only one available, and F values are calculated automatically.

Support and Availability. These functions were written in APL on an IBM 5110 microcomputer. Due to the nature of APL and an attempt to be as efficient as possible, these programs are highly compact, and therefore, difficult to follow. Detailed explanations of the algorithms of the programs with examples are available without cost from the author.

REFERENCES

- KEPPEL, G. *Design and analysis: A researcher's handbook*. Englewood Cliffs, N.J.: Prentice-Hall, 1973.
MADIGAN, S. Analysis of variance with APL/360. *Behavior Research Methods & Instrumentation*, 1976, 8, 390.

```

V UWMAOV X:CP;D;DF;HS;I;J;LV;MS;NF;NS;NV;
P;S;SP;SS;VN;XP
[1] SOURCE SS DF MS
[2] XP+(+/X)/+/X#0
[3] SS+DF+S+(J+0)*NV+/P+CP*(NF*2)+0, \NS+
(2*(NF+PLV+XP))-1
[4] VN+NF*ABCEFG
[5] HS+(+/LV)/+/+/+/X#0
[6] DUM:D+XP
[7] I+1+PSP+(0=P*(J+J+1));J)/\NF
[8] +(J=NS+1)/SQ
[9] SUM:D+/(SP*1+1-1) D
[10] +(I>1)/SUM
[11] SQ:SEJJ+HS*(+/D*2)/+/LV*SP
[12] +(J=1)/DUM
[13] MS+(SEJJ)/S*(1+NV*J)-NV*(+/P)/+/2=P
+(P)*P*J;J)/DFEJ+X/LV*CP+PJ;J)/\NFJ-1
[14] B+VNECP; 14 4 5 0 14 4 +SEJJ,DFEJ,MS
[15] +(J=NS)/DUM
[16] B+VN,'S'; 14 4 5 0 14 4 +SEJJ,DFEJ,
MS+(SEJJ)/+/X*2)/+/((+/X)*2)/+/X#0
+DFEJ+1)/+/X#0)-X/LV
[17] THE F VALUES ARE: WITH DF:
[18] FT:9+VNECP;J)/\NF; 8 3 +(SEJJ+DFEJ)/MS;
(' ' 3 0 +DFEJ+1);',',DFEJ;')
[19] +(1+NS)/FT
V

```

Figure 3. Unweighted means analysis of variance program.

(Accepted for publication June 19, 1979.)