ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
XIV, 4 (1975)

ALGORITHM 39

ANNA BARTKOWIAK (Wrocław)

EVALUATION OF CORRECTED SUMS OF SQUARES FOR ANALYSIS OF VARIANCE IN A FACTORIAL DESIGN WITH n FACTORS

1. Procedure declaration. Given a factorial design with n factors, procedure *squares* finds 2^n different sums of squares, each with appropriate degrees of freedom, nedded for testing hypotheses in the linear model of analysis of variance.

Data:

n — number of factors;
$f[1:n]$ — factor levels;
$a[0:\prod_{i=1}^{n}(f[i]+1)-1]$ — array containing data scores and all marginal means calculated, for example, by procedure *means* (see [1]);
setup — identifier of the procedure setting local integer arrays m, fl described as follows:

$$fl[i] = f[i] \text{ for } i = 1, \ldots, n, \quad m[n] = 1,$$

$$m[i] = \prod_{j=i+1}^{n}(f[j]+1) \quad \text{for } i = 1, \ldots, n-1.$$

The heading of *setup* should be

procedure *setup*(n, f, fl, m);
value n; **integer** n; **integer array** f, fl, m;

address — identifier of the procedure calculating the address of every marginal mean placed in the array a.

The heading of *address* should be

integer procedure *address*(n, f, fl, m);
integer n; **integer array** f, fl, h;

The value of *address* is set as follows:

$$address = \sum_{i=1}^{n}(f[i]-fl[i]) \times m[i].$$

Results:

$df[0:2^n-1]$ — degrees of freedom;

$ss[0:2^n]$ — corrected sums of squares; the last element $ss[2^n]$ represents the total uncorrected sum of squares of all data scores included in the array a.

```
procedure squares(n,f,a,df,ss,setup,address);
 value n;
 integer n;
 integer array f,df;
 array a,ss;
 procedure setup;
 integer procedure address;
 begin
  integer fk,i,j,k,l,mk,p,s,s1,size;
  real x,y;
  integer array b,fl,m[1:n];
  setup(n,f,fl,m);
  l:=2↑n-1;
  s:=size:=f[n];
  for k:=n-1 step -1 until 1 do
   begin
    fk:=f[k];
    size:=size×fk;
    s:=s+fk×m[k]
   end k;
  ss[0]:=a[s]×a[s]×size;
  df[0]:=1;
  for j:=1 step 1 until l do
   begin
    k:=j;
    for i:=n step -1 until 1 do
     begin
      if (-1)↑k<0
       then
```

```
          begin
           b[i]:=1;
           fl[i]:=f[i]
          end(-1)↑k lt 0
          else b[i]:=fl[i]:=0;
         k:=k+2
        end i;
       s:=address(n,f,fl,m);
       k:=n+1;
       p:=0;
       x:=.0;
   et: k:=k-1;
       if k=0
        then go to fin;
       if b[k]=0
        then go to et;
       mk:=m[k];
       fk:=f[k];
       s1:=s;
       for i:=1 step 1 until fk do
        begin
         y:=a[s1];
         x:=x+y×y;
         s1:=s1+mk
        end i;
       p:=p+fk;
   et1:k:=k-1;
       if k=0
        then go to fin;
       if b[k]=0
```

```
          then go to et1;
         s:=s+m[k];
         if fl[k]>1
          then
          begin
           fl[k]:=fl[k]-1;
           k:=n+1;
           go to et
          end fl[k] gt 0;
         fk:=fl[k]:=f[k];
         s:=s-fk×m[k];
         go to et1;
fin:ss[j]:=x×size/p;
         df[j]:=p
        end j;
      ss[l+1]:=ss[l];
      p:=l:=l+1;
      for k:=n step -1 until 1 do
       begin
         s:=p:=p÷2;
diff:
         for j:=1 step 1 until p do
          begin
           df[s]:=df[s]-df[s-p];
           ss[s]:=ss[s]-ss[s-p];
           s:=s+1
          end j;
         s:=s+p;
         if s<l
          then go to diff
          end k;
        end squares
```

2. Method used. The mathematical model for analysis of variance of a general factorial design with n factors can be found, for example, in [3]. The way of the calculation will be shown by an example with $n = 3$ factors A, B, C appearing at $f[1] = p$, $f[2] = q$, $f[3] = r$ levels. The numeration of factors is from the left to the right, e.g. $x_{i.k}$ denotes a partial mean at the i-th level of factor A and at the k-th level of factor C averaged over the factor B:

$$x_{i.k} = \frac{1}{q}\sum_{l=1}^{q} x_{ilk}, \quad 1 \leqslant i \leqslant p,\ 1 \leqslant k \leqslant r.$$

The schedule of calculations is the following:

First, the raw (or uncorrected) sums of squares are calculated. We get

$$ss[0] = SS_0 = pqr(x_{...})^2 \qquad \text{with } df[0] = 1,$$

$$ss[1] = SS_C = pq\sum_{k=1}^{r}(x_{..k})^2 \qquad \text{with } df[1] = r,$$

$$ss[2] = SS_B = pr\sum_{j=1}^{q}(x_{.j.})^2 \qquad \text{with } df[2] = q,$$

$$ss[3] = SS_{BC} = p\sum_{j=1}^{q}\sum_{k=1}^{r}(x_{.jk})^2 \qquad \text{with } df[3] = qr,$$

$$ss[4] = SS_A = qr\sum_{i=1}^{p}(x_{i..})^2 \qquad \text{with } df[4] = p,$$

$$ss[5] = SS_{AC} = q\sum_{i=1}^{p}\sum_{k=1}^{r}(x_{i.k})^2 \qquad \text{with } df[5] = pr,$$

$$ss[6] = SS_{AB} = r\sum_{i=1}^{p}\sum_{j=1}^{q}(x_{ij.})^2 \qquad \text{with } df[6] = pq,$$

$$ss[7] = SS_{ABC} = \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r}(x_{ijk})^2 \qquad \text{with } df[7] = pqr.$$

Now we set $ss[8] = ss[7]$.

The order of coding various factors in the SS terms is that proposed by Yates.

Next, the corrected sums of squares are calculated. The corrections are executed in the following $k = 3$ steps:

Step 0	Step 1	Step 2
SS_0	$\vert SS_0$	$\vert SS_0$
SS_C	$\vert SS_C$	$\vert SS_C$
SS_B	$\vert SS_B$	$\Vert SS_B - SS_0$
SS_{BC}	$\vert SS_{BC}$	$\Vert SS_{BC} - SS_C$
SS_A	$\Vert SS_A - SS_0$	$\vert SS_A - SS_0$
SS_{AC}	$\Vert SS_{AC} - SS_C$	$\vert SS_{AC} - SS_C$
SS_{AB}	$\Vert SS_{AB} - SS_B$	$\Vert SS_{AB} - SS_B - SS_A + SS_0$
SS_{ABC}	$\Vert SS_{ABC} - SS_{BC}$	$\Vert SS_{ABC} - SS_{BC} - SS_{AC} + SS_C$

Step 3

$\vert SS_0$
$\Vert SS_C - SS_0$
$\vert SS_B - SS_0$
$\Vert SS_{BC} - SS_C - SS_B + SS_0$
$\vert SS_A - SS_0$
$\Vert SS_{AC} - SS_C - SS_A + SS_0$
$\vert SS_{AB} - SS_B - SS_A + SS_0$
$\Vert SS_{ABC} - SS_{BC} - SS_{AC} + SS_C - SS_{AB} + SS_B + SS_A - SS_0$

In the first step the sums of squares are corrected for the first factor (A). From each square with the factor A the corresponding square without the factor A is substracted.

In the second step the partially corrected squares are subjected to a second correction for the factor B.

In the third step the corrections for the factor C are added. The terms left unchanged in steps 1-3 are marked with a single stroke, the terms which have changed are marked with a double stroke.

3. Certification. Let us call procedure *squares* with following values:

$$n = 3, \quad f = [2, 3, 4],$$

$$\begin{aligned}
a = [&6.5, && 2.7, && 4.0, && 4.1, && 4.325,\\
&5.2, && 4.5, && 4.1, && 3.4, && 4.300,\\
&5.6, && 4.1, && 3.6, && 5.5, && 4.700,\\
&5.767, && 3.767, && 3.900, && 4.333, && 4.442,\\
&6.5, && 4.2, && 4.7, && 4.4, && 4.950,\\
&5.1, && 3.5, && 4.9, && 5.2, && 4.675,\\
&6.1, && 3.2, && 3.7, && 3.8, && 4.200,\\
&5.900, && 3.633, && 4.433, && 4.467, && 4.608,\\
&6.500, && 3.450, && 4.350, && 4.250, && 4.637,\\
&5.150, && 4.000, && 4.500, && 4.300, && 4.487,\\
&5.850, && 3.650, && 3.650, && 4.650, && 4.450,\\
&5.833, && 3.700, && 4.167, && 4.400, && 4.525].
\end{aligned}$$

The numbers with one decimal digit visualise the original data scores, the numbers with three decimal digits were computed by procedure *means*.

Results:

$df = [1, 3, 2, 6, 1, 3, 2, 6]$,

$ss = [491.415,\ 15.218,\ 0.158,\ 2.989,\ 0.167,\ 0.340,\ 1.396,\ 3.937,\ 515.62]$.

4. Additional remarks. Procedure *squares* published here gives the same results as the procedure published by Gower [2], but the new procedure is evidently much faster, as shown in Table 1. Run on the ODRA 1204 computer our procedure needs less than 1/10 of the time needed by Gower's procedure.

TABLE 1. Run times of the old and new procedures *squares* on the ODRA 1204 computer (in seconds) (n — number of factors, $f[1:n]$ — factor levels)

parameters	$n = 3$ $f[1:3]$ $= [5, 5, 5]$	$n = 4$ $f[1:4]$ $= [5, 5, 5, 5]$	$n = 5$ $f[1:5]$ $= [4, 4, 4, 4, 4]$	$n = 6$ $f[1:6]$ $= [2, 2, 3, 4, 2, 4]$
old procedure	14	83	220	229
new procedure	1	6	16	20

References

[1] A. Bartkowiak, *Calculation of all marginal means from an n-way table*, this fascicle, p. 639-645.

[2] J. C. Gower, *Analysis of variance for a factorial table, Algorithm AS 19*, Appl. Statist. 18 (1969), p. 199-202.

[3] H. Scheffé, *The analysis of variance*, Willey, New York 1959.

COMPUTING CENTRE
MATHEMATICAL INSTITUTE
UNIVERSITY OF WROCŁAW
50-384 WROCŁAW

Received on 17. 1. 1974

ALGORYTM 39

ANNA BARTKOWIAK (Wrocław)

OBLICZANIE POPRAWIONYCH SUM KWADRATÓW DO ANALIZY WARIANCJI DLA DOŚWIADCZENIA CZYNNIKOWEGO Z n CZYNNIKAMI

STRESZCZENIE

Procedura *squares* oblicza — na podstawie jednowymiarowej tablicy danych oraz średnich marginesowych — tzw. poprawione sumy kwadratów, wyrażające zmienności wywołane wpływem badanych czynników oraz ich interakcjami. Procedura ta działa w Algolu 1204 przeciętnie 10 razy szybciej niż analogiczna procedura Gowera [2].

Dane:

n — liczba czynników;

$f[1:n]$ — poziomy czynników;

$a[0:\prod_{i=1}^{n}(f[i]+1)-1]$ — tablica rzeczywista zawierająca dane oraz średnie marginesowe; tablica taka może być obliczona za pomocą procedury *means* [1];

setup — procedura obliczająca wartości pomocniczych tablic używanych przez procedurę *squares*;

address — funkcja całkowita obliczająca adres szukanego elementu danych lub średniej marginesowej.

Wyniki:

$df[0:2^n-1]$ — stopnie swobody;

$ss[0:2^n]$ — poprawione sumy kwadratów; element $ss[2^n]$ przedstawia ogólną poprawioną sumę kwadratów.