ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
XIV, 4 (1975)

ALGORITHM 38

ANNA BARTKOWIAK (Wrocław)

CALCULATION OF ALL MARGINAL MEANS FROM AN n-WAY TABLE

1. Procedure declaration. Given a factorial design with n factors, procedure *means* finds all marginal means for all levels of all factors. The resulting values are located in a one-dimensional array together with entering data scores.

Data:

n — number of factors;

$f[1:n]$ — levels of factors;

transfer — identifier of the procedure providing, in the inverse lexicographical order, subsequent data scores obtained from the factorial design; it can contain only one instruction, e.g. the instruction of reading the subsequent data scores from punched tape;

setup — identifier of the procedure which sets the auxiliary arrays g, h as follows:

```
procedure setup(n, f, g, h);
value n; integer n; integer array f, g, h;
begin integer k; h[n] := 1; g[n] := f[n];
for k := n-1 step -1 until 1 do begin
g[k] := f[k]; h[k] := (f[k+1]+1) × h[k+1] end k
end setup
```

address — identifier of the integer procedure; given a one-dimensional array a containing all marginal means the procedure *address* finds the number s such that $a[s]$ represents the marginal mean for a given set of factor levels.

The procedure *address* should be described as follows:

```
integer procedure address (n, f, fl, h);
integer n; integer array f, fl, h;
begin integer i, k; k := 0;
for i := n step -1 until 1 do
   k := k+(f[i]-fl[i]) × h[i];
```

address := *k*
end *address*

Results:

$a[0: \prod_{i=1}^{n} (f[i]+1)-1]$ — array containing data scores (brought in by procedure *transfer*) and calculated marginal means.

2. Method used. It will be explained by an example. Let x_{ijk} represent a data score from a factorial design with $n = 3$ factors, say A, B, C. The subsequent factors are numbered from the left to the right. Let us assume further that the factor A occurs at $f[1] = 2$ levels, the factor B at $f[2] = 3$ levels and the factor C at $f[3] = 4$ levels.

The index i in x_{ijk} represents the actual level of factor A, the index j — the actual level of factor B, and the index k — the actual level of factor C. A dot on the place of an index denotes averaging over this factor.

Procedure *means* acts for the example just described as follows:

I. First the elements $x_{111}, x_{112}, x_{113}, x_{114}$ brought in by procedure *transfer* are located in the array a as elements $a[0], a[1], a[2], a[3]$. They are averaged over the third factor. So the marginal mean $x_{11.}$ is obtained and located as $a[4]$.

II. Next we change the level of the second factor. The elements $x_{121}, x_{122}, x_{123}, x_{124}$ are brought in and averaged at the same time to obtain the marginal mean $x_{12.}$. They are located as $a[5], a[6], a[7], a[8], a[9]$.

III. We change once more the level of the second factor, obtain the elements $x_{131}, x_{132}, x_{133}, x_{134}, x_{13.}$ and locate them as $a[10], a[11], a[12], a[13], a[14]$.

IV. At that point all the levels of factor B are exhausted. We calculate now the averages $x_{1.1}, x_{1.2}, x_{1.3}, x_{1.4}, x_{1..}$ and locate them as $a[15], a[16], a[17], a[18], a[19]$.

Next we repeat steps I-IV for the second level of factor A to get the elements $a[20]$-$a[39]$.

Having exhausted all levels of factor B, we calculate all means averaged over the factor A. The final result, i.e. the location of data scores and marginal means in the array a, is the following (to be read in the row sequence):

x_{111}	x_{112}	x_{113}	x_{114}	$x_{11.}$
x_{121}	x_{122}	x_{123}	x_{124}	$x_{12.}$
x_{131}	x_{132}	x_{133}	x_{134}	$x_{13.}$
$x_{1.1}$	$x_{1.2}$	$x_{1.3}$	$x_{1.4}$	$x_{1..}$
x_{211}	x_{212}	x_{213}	x_{214}	$x_{21.}$
x_{221}	x_{222}	x_{223}	x_{224}	$x_{22.}$

x_{231}	x_{232}	x_{233}	x_{234}	$x_{23.}$
$x_{2.1}$	$x_{2.2}$	$x_{2.3}$	$x_{2.4}$	$x_{2..}$
$x_{.11}$	$x_{.12}$	$x_{.13}$	$x_{.14}$	$x_{.1.}$
$x_{.21}$	$x_{.22}$	$x_{.23}$	$x_{.24}$	$x_{.2.}$
$x_{.31}$	$x_{.32}$	$x_{.33}$	$x_{.34}$	$x_{.3.}$
$x_{..1}$	$x_{..2}$	$x_{..3}$	$x_{..4}$	$x_{...}$

3. Certification. Let be given the data scores for a 3-factor design with factors A, B, C at $f[1] = 2$, $f[2] = 3$, $f[3] = 4$ levels, respectively, as shown in Table 1.

TABLE 1. A record of data from a 3-factor design

		C1	*C2*	*C3*	*C4*
A1	*B1*	6.5	2.7	4.0	4.1
	B2	5.2	4.5	4.1	3.4
	B3	5.6	4.1	3.6	5.5
A2	*B1*	6.5	4.2	4.7	4.4
	B2	5.1	3.5	4.9	5.2
	B3	6.1	3.2	3.7	3.8

The data scores are stored in the row sequence in the following order: 6.5, 2.7, 4.0, 4.1, 5.2, 4.5, 4.1, 3.4, 5.6, 4.1, 3.6, 5.5, 6.5, 4.2, 4.7, 4.4, 5.1, 3.5, 4.9, 5.2, 6.1, 3.2, 3.7, 3.8.

By call of *means* we get the following results: 6.5, 2.7, 4.0, 4.1, 4.325, 5.2, 4.5, 4.1, 3.4, 4.3, 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.

These results put together in the row sequence are given in Table 2.

TABLE 2. Data with marginal means

		C	*C2*	*C3*	*C4*	*C**
A1	*B1*	6.5	2.7	4.0	4.1	4.325
	B2	5.2	4.5	4.1	3.4	4.300
	B3	5.6	4.1	3.6	5.5	4.700
	*B**	5.767	3.767	3.900	4.333	4.442
A2	*B1*	6.5	4.2	4.7	4.4	4.950
	B2	5.1	3.5	4.9	5.2	4.675
	B3	6.1	3.2	3.7	3.8	4.200
	*B**	5.900	3.633	4.433	4.467	4.608
*A**	*B1*	6.500	3.450	4.350	4.250	4.637
	B2	5.150	4.000	4.500	4.300	4.487
	B3	5.850	3.650	3.650	4.650	4.450
	*B**	5.833	3.700	4.167	4.400	4.525

```
procedure means(n,f,a,transfer,setup,address);
 value n;
 integer n;
 integer array f;
 array a;
 procedure transfer,setup;
 integer procedure address;
 begin
  integer fk,i,k,mk,m1,r,r1,s,s1,s2;
  real x,y,z;
  integer array fl,m[1:n];
  s:=0;
  setup(n,f,fl,m);
  r:=f[n];
  k:=n-1;
  fk:=r1:=f[k];
  y:=1.0/r;
data:
  x:=.0;
  for i:=1 step 1 until r do
   begin
    transfer(z);
    x:=x+z;
    a[s]:=z;
    s:=s+1;
   end i;
  a[s]:=x×y;
  s:=s+1;
  fk:=fk-1;
```

```
  if fk>0
   then go to data;
sum:
  fl[k]:=f[k];
  s1:=address(n,f,fl,m);
  fk:=f[k];
  m1:=mk:=m[k];
  z:=1.0/fk;
sum1:
  s2:=s1;
  x:=.0;
  for i:=1 step 1 until fk do
   begin
    x:=x+a[s2];
    s2:=s2+mk
   end i;
  a[s]:=x×z;
  s:=s+1;
  m1:=m1-1;
  s1:=s1+1;
  if m1>0
   then go to sum1;
  k:=k-1;
  if k≤0
   then go to fin;
  if fl[k]=1
   then go to sum;
  fl[k]:=fl[k]-1;
  k:=n-1;
  fk:=r1;
    go to data;
  fin:
   end means
```

4. Additional remarks. Procedure *means* is the starting point for calculations of variance analysis for a factorial design.

The procedure published here gives the same results as the procedure *means* published by Gower [1]. Some trial values of run times on the ODRA 1204 computer are shown in Table 3.

TABLE 3. Run times of the old and new procedure *means* realized 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	41	300	824	757
new procedure	3	19	38	22

We see that the new procedure is much faster: it needs less than 1/10 of the time needed by Gower's procedure.

Reference

[1] J. C. Gower, *Evaluation of marginal means, Algorithm AS 18*, Appl. Statist. 18 (1968), p. 197-199.

COMPUTING CENTRE
MATHEMATICAL INSTITUTE
UNIVERSITY OF WROCŁAW
50-384 WROCŁAW

Received on 17. 1. 1974

ALGORYTM 38

ANNA BARTKOWIAK (Wrocław)

OBLICZANIE WSZYSTKICH MARGINESOWYCH ŚREDNICH W TABLICY WIELODZIELCZEJ

STRESZCZENIE

Dane są wyniki doświadczenia czynnikowego z n czynnikami. Korzystając z procedury *transfer*, która dostarcza kolejnych elementów danych w odwrotnym porządku leksykograficznym czynników, procedura *means* umieszcza poszczególne elementy tablicy danych w jednowymiarowej tablicy a. Obliczane średnie są umieszczane w tablicy a na przemian z danymi w jednoznacznie określony sposób, jak to pokazano na przykładach w punktach 2 i 3. Publikowana tu procedura działa w Algolu 1204 przeciętnie 10 razy szybciej niż analogiczna procedura Gowera [1].

Procedura *means* wchodzi w zestaw procedur obliczających analizę wariancji dla doświadczenia czynnikowego.

Dane:

n — liczba czynników;

f[1:n] — poziomy czynników;

transfer — nazwa procedury dostarczającej kolejny element danych; treścią tej procedury może być np. *read(x)*, gdzie *x* jest liczbą rzeczywistą;

setup — nazwa procedury nadającej początkowe wartości tablicom pomocniczym *g, h*;

address — nazwa funkcji całkowitej, obliczającej adres elementu tablicy danych, lub średniej marginesowej określonej za pomocą tablic *f, fl, h*.

Wyniki:

$a[0: \prod_{i=1}^{n} (f[i]+1)-1]$ — tablica rzeczywista zawierająca tablicę danych i średnie marginesowe.