CALCULATION OF ALL MARGINAL MEANS FROM AN $n$-WAY TABLE

1. Procedure declaration. Given a factorial design with $n$ factors, procedure \textit{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;
\textit{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;
\textit{setup} — identifier of the procedure which sets the auxiliary arrays $g, h$ as follows:

\begin{verbatim}
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) x h[k+1]
  end k
end setup
\end{verbatim}

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

The procedure \textit{address} should be described as follows:

\begin{verbatim}
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]) x h[i];
end address
\end{verbatim}
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 \( f[I] = 2 \) 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 \( x_{11.1}, x_{11.2}, x_{11.3}, x_{11.4}, x_{11} \) 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):

\[
\begin{align*}
&x_{111} \quad x_{112} \quad x_{113} \quad x_{114} \quad x_{11} \\
&x_{121} \quad x_{122} \quad x_{123} \quad x_{124} \quad x_{12} \\
&x_{131} \quad x_{132} \quad x_{133} \quad x_{134} \quad x_{13} \\
&x_{11.1} \quad x_{11.2} \quad x_{11.3} \quad x_{11.4} \quad x_{11} \\
&x_{211} \quad x_{212} \quad x_{213} \quad x_{214} \quad x_{21} \\
&x_{221} \quad x_{222} \quad x_{223} \quad x_{224} \quad x_{22} \\
\end{align*}
\]
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**

<table>
<thead>
<tr>
<th></th>
<th>$C1$</th>
<th>$C2$</th>
<th>$C3$</th>
<th>$C4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A1$</td>
<td>$B1$</td>
<td>6.5</td>
<td>2.7</td>
<td>4.0</td>
</tr>
<tr>
<td></td>
<td>$B2$</td>
<td>5.2</td>
<td>4.5</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td>$B3$</td>
<td>5.6</td>
<td>4.1</td>
<td>3.6</td>
</tr>
<tr>
<td>$A2$</td>
<td>$B1$</td>
<td>6.5</td>
<td>4.2</td>
<td>4.7</td>
</tr>
<tr>
<td></td>
<td>$B2$</td>
<td>5.1</td>
<td>3.5</td>
<td>4.9</td>
</tr>
<tr>
<td></td>
<td>$B3$</td>
<td>6.1</td>
<td>3.2</td>
<td>3.7</td>
</tr>
</tbody>
</table>

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.


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

**TABLE 2. Data with marginal means**

<table>
<thead>
<tr>
<th></th>
<th>$C$</th>
<th>$C2$</th>
<th>$C3$</th>
<th>$C4$</th>
<th>$C*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A1$</td>
<td>$B1$</td>
<td>6.5</td>
<td>2.7</td>
<td>4.0</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td>$B2$</td>
<td>5.2</td>
<td>4.5</td>
<td>4.1</td>
<td>3.4</td>
</tr>
<tr>
<td></td>
<td>$B3$</td>
<td>5.6</td>
<td>4.1</td>
<td>3.6</td>
<td>5.5</td>
</tr>
<tr>
<td></td>
<td>$B*$</td>
<td>5.767</td>
<td>3.767</td>
<td>3.900</td>
<td>4.333</td>
</tr>
<tr>
<td>$A2$</td>
<td>$B1$</td>
<td>6.5</td>
<td>4.2</td>
<td>4.7</td>
<td>4.4</td>
</tr>
<tr>
<td></td>
<td>$B2$</td>
<td>5.1</td>
<td>3.5</td>
<td>4.9</td>
<td>5.2</td>
</tr>
<tr>
<td></td>
<td>$B3$</td>
<td>6.1</td>
<td>3.2</td>
<td>3.7</td>
<td>3.8</td>
</tr>
<tr>
<td></td>
<td>$B2$</td>
<td>5.150</td>
<td>4.000</td>
<td>4.500</td>
<td>4.300</td>
</tr>
<tr>
<td></td>
<td>$B3$</td>
<td>5.850</td>
<td>3.650</td>
<td>3.650</td>
<td>4.650</td>
</tr>
<tr>
<td></td>
<td>$B*$</td>
<td>5.833</td>
<td>3.700</td>
<td>4.167</td>
<td>4.400</td>
</tr>
</tbody>
</table>
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 f1,m[1:n];
s:=0;
setup(n,f,f1,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;
end procedure.
if $f_k > 0$
  then go to data;

sum:
  $f_1[k] = f[k]$
  $s_1 = \text{address}(n, f, f_1, m)$
  $f_k = f[k]$
  $m_1 = m_k = m[k]$
  $z = 1.0/f_k$

$sum_1$
  $s_2 = s_1$
  $x = .0$
  for $i = 1 \text{ step } 1 \text{ until } f_k \text{ do}$
    begin
      $x = x + a[s_2]$
      $s_2 = s_2 + m_k$
    end i;
  $a[s] = x \times z$
  $s = s + 1$
  $m_1 = m_1 - 1$
  $s_1 = s_1 + 1$
  if $m_1 > 0$
    then go to $sum_1$
  $k_k = k - 1$
  if $k < 0$
    then go to $fin$
  if $f_1[k] = 1$
    then go to $sum$
      $f_1[k] = f_1[k] - 1$
      $k = n - 1$
      $f_k = r_1$
    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>
<thead>
<tr>
<th>parameters</th>
<th>( n = 3 )</th>
<th>( n = 4 )</th>
<th>( n = 5 )</th>
<th>( n = 6 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>([5, 5, 5])</td>
<td>([5, 5, 5, 5])</td>
<td>([4, 4, 4, 4, 4])</td>
<td>([2, 2, 3, 4, 2, 4])</td>
<td></td>
</tr>
</tbody>
</table>

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

Reference

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 $f, f_l, h$.

Wyniki:

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