ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
XIV, 3 (1974)

ALGORITHM 35

G. TRYBUŚ (Wrocław)

CALCULATION OF MEASURES OF STOCHASTICAL DEPENDENCE

1. Procedure declaration. The procedure *Mzalhc*, which is a modified version of the procedures given in [3], calculates the values of the coefficients of stochastical dependence d^2 and δ^2 between the random variables X and Y based on the sample $\Omega = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$. These coefficients, the first of which has been proposed by Hellwig in [2] and the second one by Czerwiński in [1], are defined by

$$d^2 = \frac{1 - \sum_{i=1}^{r} \sum_{j=1}^{s} \min(m_{ij}, p_i q_j)}{1 - 1/\min(r', s')} \tag{1}$$

and

$$\delta^2 = \frac{1 - \sum_{i=1}^{r} \sum_{j=1}^{s} \min(m_{ij}, p_i q_j)}{1 - \max(\sum_{i=1}^{r} p_i^2, \sum_{j=1}^{s} q_j^2)}, \tag{2}$$

where the data are grouped in an $r \times s$ correlation table, m_{ij} being the probability of an observation falling in the class (i, j), p_i is the probability of X falling in the class i, q_j is the probability of Y falling in the class j, and r' and s' denote the numbers of non-empty rows and columns, respectively, of the correlation table.

Data:

a1, *b1* — lower and upper values of X,
a2, *b2* — lower and upper values of Y,
r — number of classes of equal length in which the interval [*a1*, *b1*] is to be divided,
s — number of classes of equal length in which the interval [*a2*, *b2*] is to be divided,
n — number of observations of the variable (X, Y),
x, y [*1* : n] — observed values of X and Y in Ω.

Results:

H — value of the coefficient of dependence d^2,
C — value of the coefficient of dependence δ^2.

```
procedure Mzalhc(a1,b1,r,a2,b2,s,n,x,y,H,C);
 value a1,b1,r,a2,b2,s,n;
 integer r,s,n;
 real a1,a2,b1,b2,H,C;
 array x,y;
 begin
  integer i,j,k,n1,r1,s1;
  integer array p[0:r-1],q[0:s-1],m[0:r-1,0:s-1];
  real d,R,Sp,Sq,pq;
  r:=r-1;
  s:=s-1;
  for i:=0 step 1 until r do
   begin
    p[i]:=0;
    for j:=0 step 1 until s do
     m[i,j]:=0
   end i;
  for j:=0 step 1 until s do
   q[j]:=0;
  n1:=0;
  H:=(r+1)/(b1-a1);
  C:=(s+1)/(b2-a2);
  for k:=1 step 1 until n do
   begin
    i:=entier((x[k]-a1)×H);
    if 0≤i∧i≤r
     then
     begin
      j:=entier((y[k]-a2)×C);
      if 0≤j∧j≤s
```

```
          then
          begin
           p[i]:=p[i]+1;
           q[j]:=q[j]+1;
           m[i,j]:=m[i,j]+1;
           n1:=n1+1
          end 0≤j≤s
        end 0≤i≤r
      end k;
    d:=1.0/n1;
    H:=Sp:=Sq:=.0;
    r1:=s1:=0;
    for j:=0 step 1 until s do
     begin
      k:=q[j];
      if k>0
       then
       begin
        R:=k×d;
        Sq:=Sq+R×R;
        s1:=s1+1
       end k>0
     end j;
    for i:=0 step 1 until r do
     begin
      k:=p[i];
      if k>0
       then
       begin
        R:=k×d;
```

```
          Sp:=Sp+R×R;
          R:=R×d;
          r1:=r1+1;
          for j:=0 step 1 until s do
           begin
            pq:=R×q[j];
            if pq>.0
             then
             begin
              C:=m[i,j]×d;
              H:=H+(if pq<C then pq else C)
             end pq>.0
           end j
         end k>0
       end i;
      C:=(1.0-H)/(1.0-(if Sp<Sq then Sq else Sp));
      H:=(1.0-H)/(1.0-1.0/(if r1<s1 then r1 else s1))
    end Mzalhc
```

2. Method used. The calculation of the coefficients of stochastical dependence (1) and (2) of the random variables X and Y is performed as follows:

1° The variation intervals $[a_1, b_1]$ and $[a_2, b_2]$ of the variables X and Y, respectively, are divided into r and s equal subintervals, respectively. Thus one obtains an $r \times s$ correlation table.

2° Every element $(x_i, y_i) \in \Omega$ is counted in one of the $r \times s$ classes and the correlation table is formed (see Table 1), where

$$P_i = \sum_{j=1}^{s} M_{ij} \quad \text{and} \quad Q_j = \sum_{i=1}^{r} M_{ij},$$

and where M_{ij} denotes the number of elements of Ω which fall in class (i, j). If any of the observations $(x_i, y_i) \in \Omega$ falls outside of the variation rectangle $(a_1, b_1) \times (a_2, b_2)$, then it is omitted.

TABLE 1

i \ j	1	2	$\dots$	s	
1	M_{11}	M_{12}	$\dots$	M_{1s}	P_1
2	M_{21}	M_{22}	$\dots$	M_{2s}	P_2
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
r	M_{r1}	M_{r2}	$\dots$	M_{rs}	P_r
	Q_1	Q_2	$\dots$	Q_s	n

TABLE 2

1	x_i	y_i	i	x_i	y_i	i	x_i	y_i
1	.470	.339	18	−.422	.157	35	−.129	−2.092
2	.142	−.351	19	−1.313	.423	36	.274	−.239
3	.471	.121	20	−.627	.004	37	.314	−.696
4	1.440	.430	21	−.242	−.626	38	.986	1.547
5	−.829	−2.105	22	−.161	−1.036	39	.162	.568
6	−.340	1.410	23	−.514	.009	40	−1.412	−1.018
7	−.222	.479	24	−.555	1.123	41	1.885	.093
8	−1.730	−.440	25	−1.662	−.432	42	.549	−.652
9	.759	.958	26	−.340	.297	43	2.952	2.387
10	.739	1.852	27	.783	−.400	44	−1.662	−.313
11	.583	.418	28	.952	1.726	45	−.170	−.620
12	.536	1.483	29	−.710	.725	46	.423	1.857
13	−.281	−1.903	30	−.209	.338	47	.989	1.930
14	1.128	−.575	31	.328	−1.354	48	.273	.058
15	−.364	−1.187	32	−1.354	−1.470	49	.898	−1.188
16	−.014	.185	33	−.653	−.283	50	−.642	−.896
17	.802	1.804	34	.984	.271			

3° The quantities $m_{ij} = M_{ij}/n'$, $p_i = P_i/n'$ and $q_j = Q_j/n'$ $(i = 1, 2, \dots, r;\ j = 1, 2, \dots, s)$, where

$$n' = \sum_{i=1}^{r} P_i = \sum_{j=1}^{s} Q_j \qquad (n' \leqslant n),$$

are calculated and, afterwards, they are substituted into formulas (1) and (2).

3. Example. For the data given in Table 2 and for $a_1 = -3$, $b_1 = 3$, $r = 10$, $a_2 = -2.5$, $b_2 = 2.5$, $s = 10$, the following values have been obtained: $H = 0.430$ and $C = 0.471$.

References

[1] Z. Czerwiński, *O mierze zależności stochastycznej* (*On the measure of stochastical dependence*), Przegl. Statyst. 2 (1970), p. 133-146.

[2] Z. Hellwig, *On the measurement of stochastical dependence*, Zastosow. Matem. 10 (1969), p. 233-247.

[3] E. Trybuś, *Metody porządkowania zbiorów skończonych* (*Methods of ordering of finite sets*), Doctoral thesis, Graduate School of Economics, Wrocław 1972.

INSTITUTE OF STATISTICS AND ECONOMIC ACCOUNTING
GRADUATE SCHOOL OF ECONOMICS, WROCŁAW

Received on 23. 5. 1973

ALGORYTM 35

G. TRYBUŚ (Wrocław)

OBLICZENIE MIAR ZALEŻNOŚCI STOCHASTYCZNEJ

STRESZCZENIE

Procedura *Mzalhc*, która jest zmienioną wersją procedur zamieszczonych w [3], służy do obliczenia – dla zmiennych losowych X i Y – współczynnika zależności stochastycznej (1), zaproponowanego przez Hellwiga w [2], oraz współczynnika (2), zaproponowanego przez Czerwińskiego w [1] i będącego odmianą współczynnika (1).

Dane:

a1, *b1* – przedział zmienności dla X,
a2, *b2* – przedział zmienności dla Y,
r – liczba klas równej długości, na jaką należy podzielić przedział [*a1*, *b1*],
s – liczba klas równej długości, na jaką należy podzielić przedział [*a2*, *b2*],
n – liczba realizacji zmiennej losowej (X, Y),
x,*y* [*1* : *n*] – zaobserwowane wartości zmiennych X i Y.–

Wyniki:

H – wartość współczynnika zależności stochastycznej d^2,
C – wartość współczynnika zależności stochastycznej δ^2.
