

P 88

**Leopold Zlobec, Mladen Varga i Konstantin  
Momirović**

Računski centar Instituta za kineziologiju i  
Sveučilišni računski centar, Zagreb

**CONTAB — PROGRAM ZA ANALIZU NORMAL-  
NIH I ORDINALNIH VARIJABLI**

## CONTAB — THE PROGRAM FOR ANALYSIS OF NOMINAL AND ORDINAL VARIABLES

The program for analysing nominal and ordinal variables — Contab and the way of its usage is described.

CONTAB can form contingency tables for all intersections of variables from a data matrix or only those which are specially determined. Contingency tables can be done as well for one, two or three control variables. In that case CONTAB forms contingency tables for every category of control variables, and for the intersections of these categories.

Besides marginal and intersectional frequencies and marginal and intersectional probabilities and marginal conditional probabilities and usual  $\chi^2$  test, CONTAB calculates entropies of marginal variables and the entropy of contingency table, as well as conditional entropies of every marginal variable if the other is treated as the independent variable.

For nominal variables CONTAB calculates the following coefficients of association: Pearson's contingency coefficient, Kramer's coefficient of association, Garner's coefficients of relative one direction and both direction information flow and W. Pearson's estimations of symmetric and asymmetric correlation coefficients.

For ordinal variables CONTAB calculates, besides cited coefficients, still Spearman's rang coefficient, Kendal's rang coefficient for grouped data, Goodman-Kruskal's rang correlation coefficient and Somer's asymmetric and symmetric rang correlation coefficients.

Number of subjects for CONTAB is practically unlimited. Number of variables is limited to 100, and number of categories also to 100.

## CONTAB — ПРОГРАММА ДЛЯ АНАЛИЗА НОМИНАЛЬНЫХ И ОРДИНАЛЬНЫХ ПЕРЕМЕННЫХ

В настоящей работе описывается программа для анализа номинальных и ординальных переменных и способ ее использования. На основании номинальных и ординальных переменных CONTAB образует контингентные таблицы и вычисляет различные виды коэффициентов ассоциации. Число испытуемых для этого анализа является безграничным, а максимальное число переменных и категорий — 100.

## 1. UVOD

CONTAB je dio programskog sistema, razvijenog u Računskom centru Instituta za kineziologiju i Sveučilišnom računskom centru u Zagrebu, namijenjenog analizi nominalnih i ordinalnih varijabli. U mnogim sociolozijskim, i nekim kineziolozijskim, psiholozijskim, ekonomskim, kriminolozijskim, medicinskim i pedagoškim istraživanjima neophodno je analizirati, ili samo, nominalne varijable, ili tip varijabli koji se sastoji od uređenih kategorija.

CONTAB formira kontingencijske tabele za sve intersekcije svih varijabli iz jedne matrice podataka, ili samo za one intersekcije koje se posebno odrede. CONTAB to može učiniti i za jednu, dvije ili tri kontrolne varijable. U tom slučaju CONTAB formira kontingencijske tabele za svaku kategoriju kontrolnih varijabli, i za intersekcije tih kategorija.

Za svaku marginalnu varijablu CONTAB izračunava učestalosti u kategorijama te varijable i procjene vjerojatnosti da neki entitet pripada tim kategorijama.

Za ćelije dobijene ukrštanjem kategorija marginalnih varijabli CONTAB izračunava učestalost u svakoj ćeliji, vjerojatnost da neki entitet pripada toj ćeliji i uvjetne vjerojatnosti obzirom na pripadanje kategorijama marginalnih varijabli.

Pod hipotezom da su marginalne varijable stohastički nezavisne CONTAB izračunava teoretsku učestalost u ćelijama kontingencijske tabele i testira tu hipotezu  $\chi^2$  — testom.

CONTAB zatim izračunava entropije marginalnih varijabli i entropiju kontingencijske tabele, kao i uvjetne entropije svake marginalne varijable ako je druga tretirana kao nezavisna.

Za nominalne varijable CONTAB izračunava slijedeće koeficijente asocijacije: Paersonov koeficijent kontingencije, Kramerov koeficijent asocijacije, Garnerove koeficijente relativnog jednosmjernog i dvosmjernog protoka informacija, W. Paersonove procjene koeficijenta korelacije između nominalnih varijabli simetrične i asimetrične), Goodman-Kruskalove indekse prediktivne asocijacije i testira značajnost koeficijenata asocijacije na temelju omjera vjerodostojnosti.

Ako se varijable mogu tretirati kao ordinalne CONTAB izračunava, na zahtjev, i slijedeće koeficijente korelacije: Spearmanov produkt moment koeficijent korelacije ordinalnih varijabli, Kendallov koeficijent rang korelacije za grupirane podatke, Goodman-Kruskalov koeficijent rang korelacije i Somersove asimetrične i simetrične koeficijente rang korelacija.

CONTAB također izračunava značajnost svih koeficijenata rang korelacije.

Broj entiteta koji mogu biti podvrgnuti analizi praktički nije ograničen. Broj varijabli, u postojećoj verziji programa, ograničen je na 100, ali se

na zahtjev može povećati na proizvoljno velik broj. Broj kategorija u svakoj varijabli ograničen je također na 100.

## 2. METODE

Neka su  $V_p$  i  $V_q$  ( $p, q = 1, \dots, M$ ) ma koje dvije od  $M$ , u pravilu, nominalnih varijabli. Neka su  $I_i$ ,

$i = 1, \dots, K_p$ , kategorije varijable  $V_p$ , i neka su  $I_j$ ,

$j = 1, \dots, K_q$ , kategorije varijable  $V_q$ . Neka je  $E_e$ ,

$e = 1, \dots, N$ , ma koji od  $N$  entiteta, koji mogu pripadati ma kojoj od kategorija varijabli  $V_p$  i  $V_q$ .

Neka je

$$P(E_e \in I_i) = \pi_i$$

vjerojatnost da neki entitet pripada kategoriji  $I_i$  varijable  $V_p$ , i neka je

$$P(E_e \in I_j) = \pi_j$$

vjerojatnost da neki entitet pripada kategoriji  $I_j$  varijable  $V_q$ .

Neka je

$$P(E_e \in I_i \cap I_j) = \pi_{ij}$$

vjerojatnost da neki entitet pripada skupu koji je definiran intersekcijom kategorije  $I_i$  varijable  $V_p$  i kategorije  $I_j$  varijable  $V_q$ .

Neka je

$$P(E_e \in I_i \cap I_j | I_i) = \pi_{ij \cdot i}$$

uvjetna vjerojatnost da neki entitet, koji pripada kategoriji  $I_i$  varijable  $V_p$  pripada i intersekciji kategorija  $I_i$  varijable  $V_p$  i  $I_j$  varijable  $V_q$ . Neka je

$$P(E_e \in I_i \cap I_j | I_j) = \pi_{ij \cdot j}$$

uvjetna vjerojatnost da neki entitet, koji pripada kategoriji  $I_j$  varijable  $V_q$ , pripada i intersekciji kategorija  $I_i$  i  $I_j$ .

Neka su, za neki konačni skup entiteta,  $F_i$  broj entiteta u kategoriji  $I_i$ ,  $F_j$  broj entiteta u kategoriji  $I_j$ , a  $F_{ij}$  broj entiteta u skupu koji je definiran intersekcijom kategorije  $I_i$  i  $I_j$ .

Program CONTAB izračunava frekvencije  $F_i$  i  $F_j$  za varijable  $V_p$  i  $V_q$ . i formira matricu frekvencija  $F$ .

Vjerojatnosti i uvjetne vjerojatnosti program CONTAB izračunava na slijedeći način:

$$\begin{aligned} \text{Est } \pi_i &= p_i = F_i / N & i &= 1, \dots, K_p \\ \text{Est } \pi_j &= p_j = F_j / N & j &= 1, \dots, K_q \\ \text{Est } \pi_{ij} &= p_{ij} = F_{ij} / N & i &= 1, \dots, K_p \\ & & j &= 1, \dots, K_q \\ \text{Est } \pi_{ij \cdot i} &= p_{ij \cdot i} = F_{ij} / F_j & i &= 1, \dots, K_p \\ & & j &= 1, \dots, K_q \\ \text{Est } \pi_{ij \cdot j} &= p_{ij \cdot j} = F_{ij} / F_i & j &= 1, \dots, K_q \\ & & i &= 1, \dots, K_p \end{aligned}$$

Ako su varijable  $V_p$  i  $V_q$  stohastički nezavisne,

$$\pi_{ij} = \pi_i * \pi_j$$

pa otuda

$$p_{ij} = p_i * p_j$$

zbog čega je, pod hipotezom  $p_{ij} = p_i * p_j$ ,

$$T_{ij} = (F_i * F_j) / N$$

očekivana frekvencija entiteta u intersekciji kategorija  $I_i$  i  $I_j$ . CONTAB formira i matricu frekvencija  $T_{ij}$ .

Ako su varijable  $V_p$  i  $V_q$  nezavisne, varijabla

$$\chi^2 = \sum_{i=1}^{K_p} \sum_{j=1}^{K_q} (F_{ij} - T_{ij})^2 / T_{ij}$$

koja je distribuirana kao  $\chi^2$  sa

$$d = (K_p - 1) (K_q - 1)$$

stupnjeva slobode, ima očekivanu vrijednost  $d$ .

CONTAB izračunava  $\chi^2$  za svako ukrštenje varijabli  $V_p$  i  $V_q$ , stupnjeve slobode  $d$ , i

$$P(\chi^2 | d) = P$$

na temelju relacije

$$F = \chi^2 / d$$

pa je sada normalna aproksimacija Snedecorove  $F$  distribucije, ako aproksimiramo  $\infty = 1000$

$$z = ((1 - 2/9000)F^{1/3} - (1 - 2/(9d))) * ((2/9000)F^{2/3} + 2/(9d))^{-1/2}$$

nakon čega je

$$P(\chi^2 | d) = 0.5 / (1 + C_1 Z + C_2 Z^2 + C_3 Z^3 + C_4 Z^4)$$

Veličina koeficijenata  $C$  definirana je nešto kasnije.

Entropija varijable  $V_p$  je

$$H_p = - \sum_{i=1}^{K_p} p_i \log_2 p_i ;$$

entropija varijable  $V_q$  je

$$H_q = - \sum_{j=1}^{K_q} p_j \log_2 p_j$$

a entropija sistema varijabli  $V_p$  i  $V_q$

$$H_{pq} = - \sum_{i=1}^{K_p} \sum_{j=1}^{K_q} p_{ij} \log_2 p_{ij}$$

Uvjetna entropija varijable  $V_p$  ako je data varijabla  $V_q$  je

$$H_{p \cdot q} = - \sum_{i=1}^{K_p} \sum_{j=1}^{K_q} p_{ij} \log_2 p_{ij \cdot j} ;$$

analogno, uvjetna entropija varijable  $V_q$  ako je poznata varijabla  $V_p$  je

$$H_{q \cdot p} = - \sum_{i=1}^{K_p} \sum_{j=1}^{K_q} p_{ij} \log_2 p_{ij \cdot i}$$

Program CONTAB izračunava marginalne, totalnu i uvjetne entropije  $H_p$ ,  $H_q$ ,  $H_{pq}$ ,  $H_{p \cdot q}$  i  $H_{q \cdot p}$ .

Između brojnih koeficijenata asocijacije nominalnih varijabli CONTAB izračunava:

(1) Pearsonov koeficijent kontingencije (Pearson, 1901)

$$C_{pq} = (\chi^2 / (N + \chi^2))^{1/2}$$

koji se unatoč poznatih nedostataka (inkomparabilnost  $C$  koeficijenta dobijenih na temelju kontingencijskih matrica različitog reda i nedosezanja gornje granične vrijednosti 1. i kod perfektne asocijacije) iz nejasnih razloga još uvijek često upotrebljava.

(2) Cràmerov koeficijent asocijacije (Hays, 1963)

$$R_{pq} = (\chi^2 / (L - 1) N)^{1/2}$$

gdje je

$$L = \min(K_p, K_q)$$

koji se još uvijek rjeđe upotrebljava od koeficijenta kontingencije, iako nema nijedan od njegovih nedostaka.

(3) Garnerove koeficijente relativnog jednosmjernog i dvosmjernog protoka informacija, tj. koeficijente relativne uvjetne i totalne nesigurnosti (Garner, 1962; Hays, 1963; Nie, Bent i Hull, 1970)

$$\rho_{p \cdot q} = (H_p - H_{p \cdot q}) / H_p$$

gdje je  $V_p$  tretirana kao logički zavisna, a  $V_q$  kao logički nezavisna varijabla;

$$\rho_{q \cdot p} = (H_q - H_{q \cdot p}) / H_q$$

gdje je  $V_q$  tretirana kao zavisna, a  $V_p$  kao nezavisna varijabla, i

$$\rho_{pq} = (H_p + H_q - H_{pq}) / (H_p + H_q)$$

koji se može tretirati kao simetrični koeficijent asocijacije između varijabli  $V_p$  i  $V_q$ .

(4) Pearsonove koeficijente procjene korelacije nominalnih varijabli (Pearson, 1966) izvedeni na temelju činjenice da se entropija ponaša kao varijanica, a uvjetna entropija kao parcijalna varijanica.

Korelacija između  $V_p$  kao zavisna, i  $V_q$  kao logički nezavisne varijable je

$$Q_{q \cdot p} = ((1 - (H_{p \cdot q} / H_p)^2)^{1/2})$$

a korelacija između  $V_q$  i  $V_p$ , gdje je prva varijabla tretirana kao logički zavisna, a druga kao logički nezavisna

$$Q_{pq} = ((1 - (H_{q \cdot p} / H_q)^2)^{1/2})$$

Analogno jednom od mnogih načina za definiciju Paerson-Bravaisova koeficijenta korelacije,

$$Q_{qp} = (1 - H_r^2)^{1/2}$$

gdje je

$$H_r = (H_{p \cdot q} | H_p + H_{q \cdot p} | H_q) / 2$$

simetrična mjera korelacije između varijabli  $V_p$  i  $V_q$ .

Za koeficijente asocijacije, osobito ako  $N$  nije suviše veliki, testovi hipoteze  $\pi_{ij} = \pi_i \cdot \pi_j$ , utemeljeni na omjeru vjerojatnosti, superiorniji su od Pearsonova  $\chi^2$  testa\*. Hipoteza, da su varijable  $V_p$  i  $V_q$  stohastički nezavisne testirana je i na temelju funkcije

$$\chi^2 = 2N e n N + 2 H_{pq(e)} - 2 H_{p(e)} - 2 H_{q(e)}$$

gdje su

$$H_{pq(e)} = \sum_{i=1}^{K_p} \sum_{j=1}^{K_q} F_{ij} e n F_{ij}$$

$$H_{p(e)} = \sum_{i=1}^{K_p} F_i e n F_i$$

$$H_{q(e)} = \sum_{j=1}^{K_q} F_j e n F_j$$

Tako dobijeni  $\chi^2$  distributiran je kao  $\chi^2$  varijabla sa  $(K_p - 1)(K_q - 1)$  stupnjeva slobode.

CONTAB posebno računa i ovako definiran test značajnosti asocijacije varijabli  $V_p$  i  $V_q$ .

Kao koeficijent asocijacije mogu biti upotrebjeni i indeksi predikativne asocijacije (Goodman i Kruskal, 1954; Hays, 1963; Nie, Bent i Hull, 1970).

Koeficijent

$$\lambda_{p \cdot q} = \left( \sum_{j=1}^{K_q} \phi_j - \phi_i \right) / (N - \phi_i)$$

gdje je

$$\phi_{ij} = \max_i F_{ij}$$

a

$$\phi_i = \max_j F_{ij}$$

je proporcionalan redukciji vjerojatnosti pogreške procjene pripadanja kategorijama  $I_i$  ako je poznato pripadanje entiteta kategorijama  $I_j$ .

\* za veliki  $N$  i ti su testovi praktički ekvivalentni

Analogno

$$\lambda_{q \cdot p} = \left( \sum_{i=1}^{K_p} \phi_{ji} - \phi_j \right) / (N - \phi_j)$$

gdje je

$$\phi_{ji} = \max_j F_{ij}$$

a

$$\phi_j = \max_i F_{ij}$$

proporcionalan je redukciji pogreške pripadanja kategorijama varijable  $V_q$ , ako je poznato pripadanje kategorijama varijable  $V_p$ .

Simetrični koeficijent redukcije pogreške je

$$\lambda_{pq} = \left( \sum_{j=1}^{K_q} \phi_{ij} + \sum_{i=1}^{K_p} \phi_{ji} - \phi_i - \phi_j \right) / (2N - \phi_i - \phi_j)$$

CONTAB računa sva tri  $\lambda$  koeficijenta.  $\chi^2$  test za koeficijente asocijacije je i test značajnosti  $\lambda$  koeficijenta.

Ako su kategorije varijabli  $V_p$  i  $V_q$ , za sve  $p, q = 1, \dots, M$  uređene tako da vrijednosti kategorija leže na ordinalnoj skali, CONTAB, na poseban zahtjev, računa slijedeće mjere asocijacije ordinalnih varijabli:

(1) Spearmanov produkt-moment koeficijent korelacije ordinalnih varijabli

$$S_{pq} = 1 - (6 \sum_{e=1}^N d_e^2) / (N^3 - N)$$

koji je, ako su varijable ordinalne, naravno jednak koeficijentu (Guilford, 1956; Hays, 1963)

$$S_{pq} = \left( \sum_{e=1}^N x_{pe} x_{qe} - \left( \sum_{e=1}^N x_{pe} \right) \left( \sum_{e=1}^N x_{qe} \right) / N \right) *$$

$$* \left( \sum_{e=1}^N x_{pe}^2 - \left( \sum_{e=1}^N x_{pe} \right)^2 / N \right)^{-1/2} *$$

$$* \left( \sum_{e=1}^N x_{qe}^2 - \left( \sum_{e=1}^N x_{qe} \right)^2 / N \right)^{-1/2}$$

jer je

$$d_e = x_{pe} - x_{qe}$$

a za ordinalne varijable su

$$\sum_{e=1}^N x_{pe}, \sum_{e=1}^N x_{qe}, \sum_{e=1}^N x_{pe}^2 \text{ i } \sum_{e=1}^N x_{qe}^2$$

funkcije od  $N$ .

Za grupirane podatke,

$$S_{pq} = \left( N^{-1} \sum_{j=1}^{K_q} \sum_{i=1}^{K_p} F_{ij} i j - \left( \sum_{j=1}^{K_q} F_j j \right) \right)$$

$$\left( \sum_{i=1}^{K_p} F_i - i \right) (N-2)^*$$

$$* (N-1 \sum_{i=1}^{K_q} F_j - j^2 - (N-1 \sum_{i=1}^{K_q} F_j - j))^2)^{-1/2} *$$

$$* (N-1 \sum_{i=1}^{K_p} F_i - i^2 - (N-1 \sum_{i=1}^{K_d} F_i - i))^2)^{-1/2}$$

a tako CONTAB i računa  $S_{pq}$ .

Značajnost koeficijenta  $S_{pq}$  računa CONTAB na temelju relacije

$$F_{spq} = S_{pq}^2 (N-2) / (1 - S_{pq}^2)$$

$F_{spq}$  je pod hipotezom  $S_{pq} = 0$  distribuiran kao Snedecorov F sa 1, N-2 stupnjeva slobode.

Vjerojatnost da je u populaciji,  $S_{pq} = 0$ , izračunava CONTAB na temelju normalne aproksimacije F distribucije (Veldman, 1967)

$$z_{pq} = (1 - 2/9 (N-2)) F_{spq}^{1/2} - 7/9) / ((2/9 (N-2)) F_{spq}^{2/3} + 2/9)^{1/2}$$

pa je vjerojatnost P ( $F_{spq} | 1, N-2$ )

$$P = 0.5 / (1 + c_1 z_{pq} + c_2 z_{pq}^2 + c_3 z_{pq}^3 + c_4 z_{pq}^4)$$

gdje su koeficijenti c

$$c_1 = .196854$$

$$c_2 = .115194$$

$$c_3 = .000344$$

$$c_4 = .019527$$

CONTAB upotrebljava u tu svrhu Veldmanovu subrutinu PRBF; istu subrutinu upotrebljava i za određivanje vjerojatnosti Pearsonova  $\chi^2$  i  $\chi^2$  utemeljenog na omjeru vjerodostojnosti za testiranje značajnosti asocijacijskih koeficijenata.

(2) Kendalov  $T_c$  koeficijent rang korelacije za grupirane podatke, gdje broj grupa, tj.  $K_p$  i  $K_q$  može biti nejednak (Kendal, 1955; Blalock, 1972; Nie, Bent i Hull, 1970).

Neka je

$$V = \sum_{k=1}^{K_p-1} \sum_{e=1}^{K_q-1} F_{ke} \sum_{i=k+1}^{K_p} \sum_{j=e+1}^{K_q} F_{ij}$$

$$W = \sum_{k=1}^{K_p-1} \sum_{e=2}^{K_q-1} F_{zk} \sum_{i=2}^{K_p} \sum_{j=1}^{e-1} F_{ij}$$

Sada je  $T_c$

$$T_c = (2L(V-W)) / (N^2(L-1))$$

gdje je, kao i ranije,

$$L = \min(K_p, K_q)$$

(3) Goodman-Kruskalov  $\gamma$  koeficijent (Goodman i Kruskal, 1954; Blalock, 1972; Nie, Bent i Hull, 1970; Hays, 1963). Ovaj se koeficijent, definiran relacijom

$$\gamma = (V - W) / (V + W)$$

interpretira nešto lakše nego  $T_c$ .

(4) Somersove asimetrične i simetrične D koeficijente (Somers, 1962; Blalock, 1972; Nie, Bent i Hull, 1970).

Ako se  $V_p$  tretira kao logički zavisan od  $V_q$

$$D_{p \cdot q} = (V - W) / ((N^2 - \sum_{j=1}^{K_p} j F_j^2) / 2)$$

a ako se  $V_q$  tretira kao logička zavisna od  $V_p$

$$D_{q \cdot p} = (2(V - W)) / (N^2 - \sum_{i=1}^{K_p} i F_i^2)$$

Simetrična mjera asocijacije između  $V_p$  i  $V_q$  je

$$D_{pq} = (2(V - W)) / ((N^2 - \sum_{j=1}^{K_q} j F_j^2) +$$

$$+ (N^2 - \sum_{i=1}^{K_p} i F_i^2)) / 2)$$

Značajnost sva tri koeficijenta utemeljenih na razlici između V i W ( $T_c$ ,  $\gamma$ ,  $D_{p \cdot q}$ ,  $D_{q \cdot p}$ ,  $D_{pq}$ ) testirana je na slijedeći način:

$$z = (V - W) / \sigma_{v-w}$$

gdje je

$$\sigma_{v-w}^2 = \alpha + \beta + \delta$$

gdje je

$$\alpha = 18^{-1} (N(N-1)(2N+5) - \sum_{i=1}^{K_p} F_i$$

$$(F_i - 1)(2F_i + 5) - \sum_{j=1}^{K_q} F_j (F_j - 1)(2F_j + 5))$$

$$\beta = (((\sum_{i=1}^{K_p} F_i (F_i - 1)(F_i - 2)) (\sum_{j=1}^{K_q} F_j (F_j - 1)(F_j - 2))) * ((9N(N-1)(N-2))^{-1})$$

$$\delta = (((\sum_{i=1}^{K_p} F_i (F_i - 1)) (\sum_{j=1}^{K_q} F_j (F_j - 1))) * ((2N(N-1))^{-1})$$

Vjerojatnost P (z) CONTAB izračunava subrutinom PRBF, jednako kao i za ostale mjere asocijacije.

Sve operacije CONTAB može izvršiti i uz kontrolu neke varijable  $V_1$ . U tom slučaju sve dobijene entropije i koeficijenti asocijacije su, u stvari, parcijalne entropije i parcijalni koeficijenti asocijacije za skupine entiteta koji pripadaju svakoj od kategorija varijable  $V_1$ .

Ako je  $N$  dovoljno veliki, od CONTABA se može zahtijevati da analizira kontingencije varijabli  $V_p$  i  $V_q$  uz kontrolu dvije varijable  $V_2$  i  $V_3$ . U tom slučaju sve entropije i svi koeficijenti asocijacije su, u stvari, parcijalizirani za skupove entiteta koji pripadaju intersekcijama kategorija varijabli  $V_2$  i  $V_3$ .

### 3. OPIS PROGRAMA I NAČIN NJEGOVA AKTIVIRANJA

Nakon što se u kontrolnim naredbama inicijalizira RUN (JOB), te pozove programska datoteka u kojoj se nalazi program, slijedi naredba za izvršenje programa ( $\partial$ XQT), a zatim podaci za CONTAB.

Iza naredbe kojom se traži izvršenje programa dolaze podaci: PROBLEM kartica kojom se definiraju glavni parametri obrade, TEKST kartica za unošenje teksta, format kartice, kartice za imenovanje varijabli i kategorija, podaci koji se obrađuju i TAB kartice kojima se određuju intersekcije pojedinih varijabli.

Slijed ovih podataka je ovakav:

#### 1. Problemska kartica

PROBLEM, NV= $n_1$ , FORMAT= $n_2$ , ULAZ= $n_3$ , KT= $n_4$ , J= $n_5$ .

Problemom karticom definiraju se osnovni parametri obrade. Riječ PROBLEM je obavezna te deklarira karticu kao problemsku. Svi parametri se ne moraju uvijek navesti (ne moraju se navesti ukoliko su jednaki nuli), a međusobni poredak nije bitan. Vrijednosti parametra  $n_1$ — $n_5$  su cjelobrojne vrijednosti. Parametar NV= $n_1$  određuje broj varijabli u toku obrade, FORMAT= $n_2$  broj format kartica koje će se učitati, ULAZ= $n_3$  ulaznu jedinicu (ili datoteku) odakle će se čitati podaci za obradu, KT= $n_4$  kriterij teoretske frekvencije, te J= $n_5$ . Ukoliko parametar J nije naveden smatra se da su varijable nominalne, te se kao izlaz dobivaju kontingencijske table s koeficijentima opisanim u poglavlju 2. Za vrijednost parametra J=1 smatra se da varijable imaju uređene kategorije, te se u izlaznoj listi dobivaju još i slijedeće mjere asocijacije: (1) Spearmanov produkt momenat koeficijent korelacije, (2) Kendallov  $T_c$  koeficijent rang korelacije, (3) Goodman-Kruskalov  $\gamma$  koeficijent i (4) Somersovi asimetrični i simetrični D koeficijenti.

#### 2. Kartica s tekstom

Ova kartica nije obavezna. Tekst koji se unosi preko ove kartice štampa se kao naslov na svakoj stranici izlazne liste. TEKST = tekst.

#### 3. Kartice s formatom

Broj ovih kartica je određen parametrom FORMAT s problemske kartice. Kartica s formatom se upisuje po pravlima koja vrijede za format naredbe u FORTRAN-u. Budući da su podaci cjelobrojne vrijednosti mogu se koristiti I, T i X kodovi za format. Ukoliko je parametar  $n_2$  jednak nuli ove kartice se ne učitavaju, a čitanje podataka je neformatizirano.

#### 4. Kartice s imenima varijabli i kategorija

Svaka varijabla mora imati karticu kojom se definiraju ime varijable i imena kategorija. Poretom ovih kartica određen je redni broj varijabli, te treba pripaziti da je redni broj u skladu s FORMAT karticom.

$(m_1 = \text{Ime}_{m_1}, m_1 + 1 = \text{Ime}_{m_1 + 1}, \dots$   
 $\dots m_2 = \text{Ime}_{m_2})$

IMEVAR = Imev ( $m_1, m_2$ ) Ime $_{m_1}$  Ime $_{m_1 + 1}$   
 $, \dots \dots \text{Ime}_{m_2}$

IMEVAR je obavezna riječ koja se buši na početku kartice. Imev je ime varijable; to je niz od 1 — 6 alfanumeričkih ili specijalnih znakova. Brojevima  $m_1, m_1 + 1, \dots, m_2$  unutar zagrada određene su šifre kategorija unutar varijable Imev.  $m_1$  je šifra prve kategorije, a  $m_2$  šifra posljednje kategorije.

Ime $_{m_1} \dots \text{Ime}_{m_2}$  su imena kategorija. Kao i Imev ova imena su nizovi od 1 — 6 alfanumeričkih znakova. Kartice s imenima varijabli i kategorija nisu neophodne. Ukoliko nisu navedene, sistem će uzeti kao imena varijabli imena VAR $\Theta\Theta$  1, VAR $\Theta\Theta$  2 itd, a broj kategorija je ograničen maksimumom kategorija koje program dopušta.

#### 5. Kartice s podacima

Ove se kartice navode ukoliko parametar ULAZ s problemske kartice deklarira čitač kartica kao ulaznu jedinicu, a završavaju se  $\partial$ EOF. Kad se podaci nalaze u datoteci na vanjskoj memoriji, parametrom ULAZ se navode ime datoteke, te ove grupe kartica nema.

#### 6. Kartice za određivanje intersekcija

Ovim karticama se određuju zahtijevane intersekcije pojedinih varijabli i time određene kontingencijske tabele.

Format ovih kartica i primjeri:

##### a) TAB = 3

Izračunavanje učestalosti marginalne varijable 3.

##### b) TAB = 3 — 8

Izračunavanje učestalosti marginalnih varijabli 3 do 9.

##### c) TABSVE

Izračunavanje učestalosti svih varijabli.

##### d) TAB = 1, 7

Intersekcija varijabli 1 i 7.

- e) TAB = 1, 4 — 8  
Intersekcija varijable 1 s varijablom 4 do 8.
- f) TAB = 1 — 3, 4 — 8  
Intersekcija varijabli 1 do 3 s varijablama 4 do 8.
- g) TABSVE = 1, 7  
Sve moguće kombinacije intersekcija varijabli 1 do 7.
- h) TAB = 1, 2/4  
Intersekcija varijabli 1 i 2 uz kontrolu varijable 4.
- i) TAB = 1, 2/4/8  
Intersekcija varijabli 1 i 2 uz dvije kontrolne varijable: 4 i 8.
- j) TAB = 1, 2 — 3/4/8  
Intersekcije varijable 1 s varijablama od 2 do 3 uz kontrolne varijable 4 i 8.

Primjer cijelog programa (uz kontrolne naredbe na računalima iz serije UNVAC 1100):

```

@RUN
@ASG, A ime-datoteke
@XQT ime-programa
PROBLEM, NV=3, FORMAT=1, ULAZ=5, KT=1,
J=@.
TEKST = PRIMJER
(T12, 3I2)
IMEVAR=ŠKOLA (1, 4) BEZŠK, OSNOV, SRED,
FAKUL
IMEVAR=USPJEH (2,5) DOVOLJ, DOBAR, VRLO,
ODLIC
IMEVAR=TPNAS (1, 3) SEOSKO, MJES, GRAD

```

podaci

```

@EOF
TABSVE
TAB=1, 3
TAB=1, 2/3
@FIN

```

#### 4. REZIME

Opisan je program za analizu nominalnih i ordinalnih varijabli CONTAB i način njegove primjene.

CONTAB formira kontingencijske tabele za sve intersekcije svih varijabli iz jedne matrice podataka, ili samo za one intersekcije koje se posebno odrede. CONTAB to može učiniti i za jednu, dvije ili tri kontrolne varijable. U tom slučaju CONTAB formira kontingencijske tabele za svaku kategoriju kontrolnih varijabli, i za intersekcije tih kategorija.

Za svaku marginalnu varijablu CONTAB izračunava učestalost u kategorijama te varijable i pro-

cjene vjerojatnosti da neki entitet pripada tim kategorijama.

Za ćelije dobijene ukrštanjem kategorija varijabli CONTAB izračunava učestalost u svakoj ćeliji, vjerojatnost da neki entitet pripada toj ćeliji i uvjetne vjerojatnosti obzirom na pripadanje kategorijama marginalnih varijabli.

Pod hipotezom da su marginalne varijable stohastički nezavisne CONTAB izračunava teoretsku učestalost u ćelijama kontingencijske tabele i testira tu hipotezu  $\chi^2$  — testom.

CONTAB zatim izračunava entropije marginalnih varijabli i entropiju kontingencijske tabele, kao i uvjetne entropije svake marginalne varijable ako je druga tretirana kao nezavisna.

Za nominalne varijable CONTAB izračunava slijedeće koeficijente asocijacije: Pearsonov koeficijent kontingencije, Kramerov koeficijent asocijacije, Garnerove koeficijente relativnog jednosmjernog i dvosmjernog protoka informacija, W. Pearsonove procjene koeficijenta korelacije između nominalnih varijabli (simetrične i asimetrične) i testira značajnost koeficijenta asocijacije na temelju omjera vjerodostojnosti.

Ako se varijable mogu tretirati kao ordinalne, CONTAB izračunava, na zahtjev, i slijedeće koeficijente korelacije: Spearmanov produkt moment koeficijent korelacije ordinalnih varijabli, Kendallov koeficijent rang korelacije za grupirane podatke, Goodman-Kruskalov koeficijent rang korelacije i Somersove asimetrične i simetrične koeficijente rang korelacije.

CONTAB također izračunava značajnost svih koeficijenata rang korelacije.

Broj entiteta koji mogu biti podvrgnuti analizi praktički nije ograničen. Broj varijabli, u postojećoj verziji programa, ograničen je na 100, ali se na zahtjev može povećati na proizvoljno velik broj. Broj kategorija u svakoj varijabli ograničen je također na 100.

#### 5. LITERATURA

1. Blalock, H. M. Social statistics (2<sup>ed.</sup>) McGraw-Hill New York, 1972.
2. Garner, W. R. Uncertainty and structure as psychological concepts. Wiley, New York, 1962.
3. Goodman, L. A. and W. H. Kruskal Measures of association for crossclassification. Journal of American Statistical Association, 1954., 49 pp 732—764
4. Guilford, J. P. Fundamental statistics in psychology and education. McGraw-Hill, New York, 1956.
5. Hays, W. L. Statistics. Holt, Rinehart and Winston, New York, 1963.



6. Kendall, M. G.  
The advanced theory of statistics (Vol. I),  
Griffin, London, 1947.
7. Kendall, M. G.  
Rank correlation methods (2.ed.), Griffin, Lon-  
don, 1955.
8. Nie, N. N., D. H. Benet and C. H. Hull  
SPSS-Statistical package for the social scien-  
ces. McGraw-Hill, New York, 1970.
9. Pearson, K.  
On the correlation of characters not quanti-  
tatively measurable.  
Philosophical Transactions, Series A 1901, 195,  
pp. 1—47
10. Pearson, W. H.  
Estimation of a correlation coefficient from a  
uncertainty measure.  
Psychometrika, 1966, 31, pp. 421 — 433
11. Somers, R. H.  
A new asymmetric measure of association for  
ordinal variables. American Sociological Revi-  
ew. 1962, 27, pp. 799 — 811.
12. Veldman, D. J.  
Fortran programming for the behavioral sci-  
ences. Holt, Rinehart and Winston New York,  
1967.

