

Un program de validare a datelor recoltate prin chestionarul sociologic

Petruș Alexandrescu

Dan Mateescu

În etapa actuală de făurire a societății noastre socialiste multilateral dezvoltate, cind știința devine forță nemijlocită de producție, un rol deosebit de important îl capătă, în cadrul procesului complex al revoluției științifice și tehnice, utilizarea pe scară tot mai largă a tehnicii electronice de calcul, atât în procesul producției, al creșterii gradului de tehnicizare a acesteia, dar și în dezvoltarea tuturor compartimentelor științei. Științele sociale beneficiază, la rindul lor, de sprijinul tot mai susținut al calculatoarelor electronice, odată cu procesul tot mai avansat al matematizării științelor. Adoptarea unor tehnici și metode statistico-matematice ca și elaborarea unora specifice în studiul fenomenului social, au condus la cererea de programe adecvate metodologiei de cercetare a științelor sociale, în speță a sociologiei. În zilele noastre practic de neconceput ca o cercetare sociologică, care are în vedere pentru studiul unui fenomen social, o bogată bază de date statistice, ancheta de opinie etc., să se poată desfășura la nivelul exigențelor actuale ale științei, fără contribuția tot mai mare a diferitelor discipline ale matematicii. Dar problema măsurării, a analizelor cantitative implică un volum tot mai ridicat de calcule, care presupun utilizarea calculatorului electronic. Necesitatea utilizării calculatorului electronic rezidă nu numai în volumul tot mai mare de date recoltate care trebuie acumulate, sortate, validate, stocate și apoi prelucrate în scopul urmărit de cercetarea sociologică, dar și în faptul că studiul unui fenomen social atrage după sine analiza unui mare număr de variabile cu grade diferite de interdependență, de clasificare și ordonare a acestora după influențele pe care le primesc sau le exercită etc. Toate aceste conexiuni se stabilesc printr-o serie de metode, tehnici și modele matematice, mergînd pînă la modelele de simulare pe calculator prin control parametric.

Toate aceste lucruri sînt cunoscute și frecvente în cercetările sociologice.

În cele ce urmează, ne propunem să prezentăm schemele logice ale unui program de validare a informațiilor obținute prin chestionarul sociologic. Necesitatea efectuării acestui program a pornit de la faptul că în realizarea unui eșantion reprezentativ ridicat (2—3 000 subiecți), apar diferite erori (în completare, de perforare a cartelelor etc.), semnalate apoi prin „citirea directă” la calculator. Scopul urmărit a fost de a elimina această informație local greșită, și de a reține restul informației corecte cu pondere importantă în semnificația eșantionului.

În cazul în speță, testarea a fost efectuată asupra unui eșantion de 2 700 de subiecți, care a constituit obiectul de studiu al unei cercetări privind informarea în masă și activitățile politico-ideologice în rîndul populației navetiste. Cercetarea s-a desfășurat în județele Dolj, Galați și Sibiu, iar pentru realizarea eșantionului a fost stabilit un procent de circa 3% din totalul personalului celor 16 unități industriale în care ne-am desfășurat investigația.

Chestionarul aplicat populației studiate a cuprins un număr de 105 întrebări, fiecare avînd între 2 și 15 posibilități de răspuns.

Răspunsurile la întrebări au fost codificate și perforate pe cartele, fiecare chestionar avînd afectate trei cartele cu următoarea structură: cartelele de tip 1, în care coloanele 1—4 conțin numărul de ordine, coloanele 5—78 conțin informația utilă, coloanele 79—80 sînt libere; cartelele de tip 2: coloanele 1—4 conțin numărul de ordine iar restul cartelei informația utilă; cartelele de tip 3: coloanele 1—4 conțin numărul de ordine, coloanele 5—43 conțin informația utilă, restul cartelei este liber. În fiecare cîmp de informație utilă a fost perforat răspunsul codificat la o întrebare.

Cu aceste date s-a creat un fișier în organizarea secvențială pe bandă magnetică.

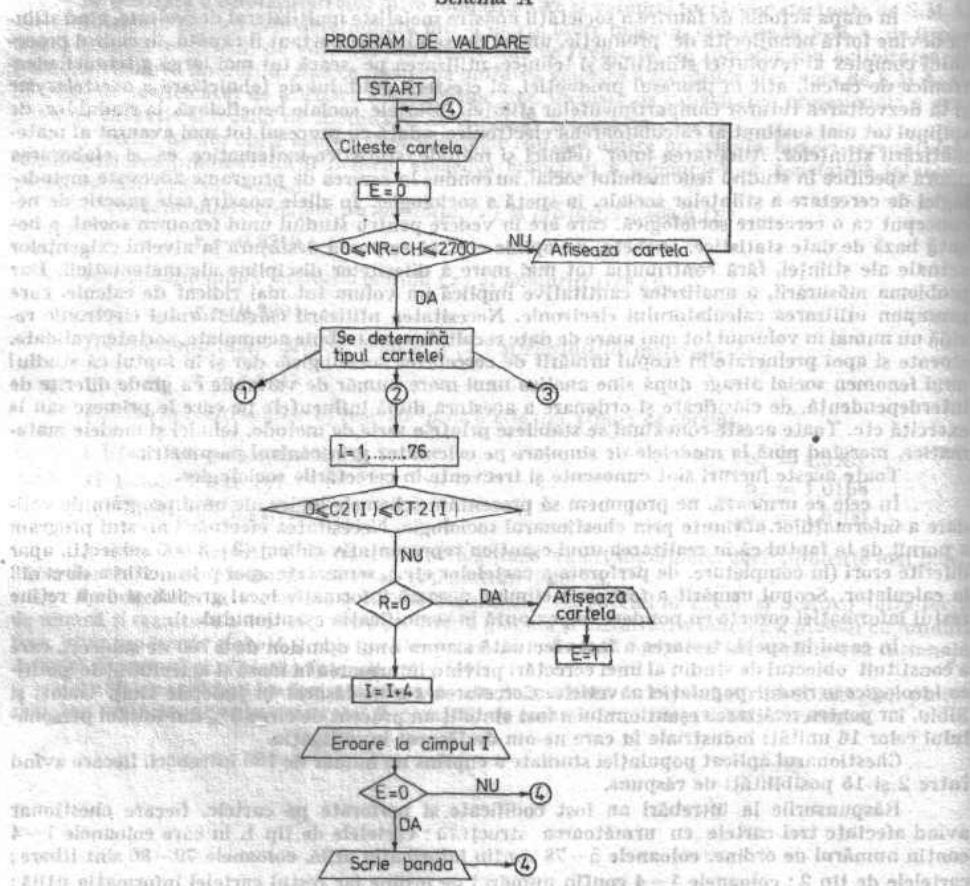
O primă operație necesară pentru crearea fișierului a constat în validarea setului de date, operație ce s-a desfășurat în mai multe faze.

Prin validare s-au obținut următoarele: cimpurile care conțineau numărul de ordine și răspunsurile la întrebări să fie numerice, apoi s-au testat cimp cu cimp datele inițiale pentru a stabili dacă cifrele corespunzătoare răspunsurilor se află între limitele impuse de structura chestionarului. Această operație a permis, totodată, corectarea unui număr de aproximativ 250 de cartele care au fost invalidate inițial din cauza unor depășiri de limite. Ceea ce a fost foarte important în această fază a fost observarea unor erori grupate care se refereau la categorii sociale restrinse din punct de vedere numeric, în acest caz renunțarea la chestionarele rău întocmite ar fi dus la erori destul de mari asupra reprezentativității straturilor.

Cartelele invalidate din cauza unor erori negrupate au fost recuperate prin introducerea unui cod special în cimpul eronat. În acest mod restul informației a putut fi folosit.

În prima fază cartelele invalide au fost afișate la imprimantă cu specificații asupra tuturor cimpurilor eronate, cartelele corecte au fost trecute în fișier. Schema logică a acestui program este prezentată în schema A.

Schema A



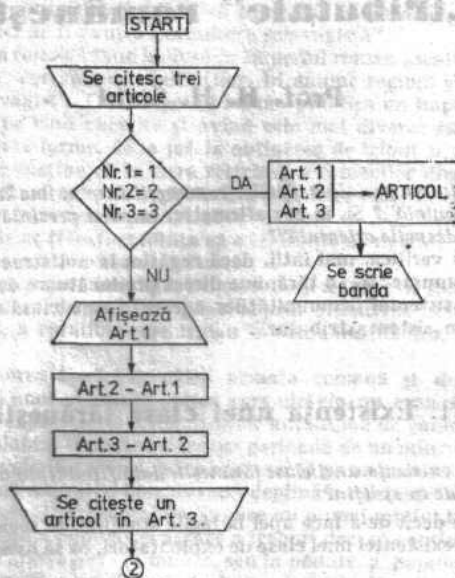
În a doua fază au fost corectate cartelele invalide și au fost adăugate la sfârșitul fișierului. Fiecărei imagine de cartelă din fișierul pe bandă i s-a atașat încă un cimp între cimpurile 4 și 5 ale cartelei; acest cimp conținea cifra 1, 2 sau 3 corespunzătoare tipului cartelei.

Datorită faptului că datele unui subiect trebuiau corelate s-a impus operația de „alipire” a cartelelor corespunzătoare aceluiași subiect și crearea, în acest mod, a unei fișe personale pentru fiecare respondent. Pentru aceasta a fost sortat fișierul după cheile formate din primele 5 cimpuri ale fiecărui articol.

În final fișierul, astfel obținut, a fost testat articol cu articol pentru a se controla succesiunea a exact trei cartele de tip 1, 2, 3 pentru fiecare număr de ordine.

O variantă a acestui program prevede și înregistrarea fiecărui cod, care inițial se afla pe un octet, pe o lungime de un cuvânt pentru ca fișierul să poată fi exploatat și în FORTRAN. Schema logică a acestui program este indicată în schema B.

Schema B



Programul folosește ca date de intrare banda sortată avînd articole de trei feluri: Art. 1, Art. 2, Art. 3, corespunzătoare celor trei tipuri de cartele. Se controlează secvența, și, în cazul în care este cea dorită, articolele se trec în ARTICOL care se scrie pe banda de ieșire.

Programul utilizează ca date de intrare cartelele obținute în urma codificării chestionarelor precum și trei cartele de test. În prima fază cartelele invalide sînt ignorate. În urma acestor faze se poate face, sau corectarea cartelei pentru confruntare cu chestionarul respectiv sau numai ignorarea codului respectiv, reținîndu-se restul informației. Secvențele 1 și 3 sînt analoge cu 2.