

OBRADA ANKETNIH PODATAKA NA MIKRORAČUNARIMA - SISTEM ISSA

Anketna istraživanja u uslovima savremene tehnologije

Prava informacija u pravom trenutku bitan je faktor donošenja odluka u svim sferama života. Zato se puno pažnje posvećuje razvoju metoda prikupljanja, obrade i diseminaciji podataka.

Anketnim istraživanjima prikupljaju se podaci o jedinicama konačne populacije. Na primer, o troškovima domaćinstva, o fertilitetu žena, o prihodima i poslovanju preduzeća, o prinosu žitarica poljoprivrednih gazdinstava, o javnom mnjenju itd. Pri tome se, uz poštovanje unapred postavljenih vremenskih rokova i materijalnih troškova, teži da prikupljeni podaci budu što kvalitetniji. Pod kvalitetom podataka podrazumeva se njihova tačnost i korisnost (odnosno koliko prikupljeni podaci daju odgovore na predmet istraživanja). Za anketno istraživanje sa jasno postavljenim ciljevima, kvalitet prikupljenih podataka zavisi od toga kako su isplanirane i sprovedene pojedine faze anketnog istraživanja, među kojima izdvajamo:

- Kreiranje upitnika
- Priprema okvira za izbor jedinica
- Izbor tipa uzorka i sistema ocenjivanja
- Izbor jedinica uzorka prema planu uzorka
- Obuka anketara i kontrolora
- Prikupljanje podataka na terenu
- Mašinska obrada: programi za unos, kontrolu, korekciju, imputaciju i tabeliranje podataka.
- Priprema anketnih izveštaja

Savremena tehnologija uticala je na sve faze anketnih istraživanja pružajući mogućnosti da se one integrišu, paralelno i interaktivno odvijaju na decentralizovan način. Uz optimalnu primenu savremene tehnologije dolazi se do znatno kvalitetnijih rezultata anketnih istraživanja uz velike uštede materijalnih potencijala i vremena potrebnog za izvršenje pojedinih koraka. Osim što se time obezbeđuje aktuelnost podataka, uštede u vremenu i materijalnim sredstvima mogu se usmeriti na faze planiranja i pripreme tako da se optimalno iskoriste i stručna lica angažovana u istraživanjima.

Savremena tehnologija je naročito izmenila način obrade i prikupljanja podataka. Tako se, osim tradicionalnog metoda, anketiranja putem intervjua uz pomoć olovke i upitnika na listu papira (PAPI), podaci anketnih istraživanja prikupljaju putem telefona uz pomoć računara (CATI), i uz pomoć *laptop* računara (CAPI). Poslednjim dvema metodama podaci dobijeni od ispitanika direktno se smeštaju u memoriju računara. Na taj način su integrisane faze prikupljanja i obrade.

U slučaju tradicionalnog metoda prikupljanja podataka, kada je potrebno da se podaci sa papira unesu u memoriju računara, proces obrade je takođe izmenjen. Umesto da unos obavljaju profesionalni daktilografi neinteligentnim programima za unos (metod "*heads down*"), sve više se, uz pomoć inteligentnog interaktivnog programa, unos prepušta stručnim licima za oblast istraživanja. Mnogi poslovi kontrole i korekcije koji su se radili isključivo u *batch* obradi (obradi u paketu) prenose se na fazu unosa.

U našim uslovima PAPI i CATI anketiranja se mnogo češće sprovode od CAPI anketiranja za koje su potrebni *laptop* računari. Istraživanja u svetu (Baker i sar; 1995) su pokazala da, bez obzira na veću startnu cenu CAPI od PAPI istraživanja, ima mnogo razloga koji idu u prilog uvođenja CAPI anketiranja. Na primer, pri CAPI anketiranju, niža je stopa ispitanika koji ne daju odgovore (*nonresponse rate*) na pojedina pitanja, veća je spremnost ispitanika da daju odgovore na osetljiva pitanja i u proseku je kraće vreme anketiranja za oko 20%. CAPI anketiranje daje bržu i kvalitetniju obradu podataka po nižoj ceni. Uz dobru obuku anketara i padom cena portabilnih računara ovaj metod će postajati sve primamljiviji.

Obrada anketnih podataka sve više se prenosi sa velikih mašina računskih centara na mikroracunare anketnih službi. Razvijeni su softverski paketi specijalno namenjeni obradi podataka na mikroracunarima. Među njima je veoma poznat Blaise sistem koji je razvio tim stručnjaka Holandskog centralnog statističkog biroa. Blaise je mrežni softver koji je doprineo modernizaciji statističkih procesa statističkih službi u Holandiji ali i širom sveta. Ubraja se među vodeći softver te vrste. Najnovija verzija, Blaise III, dostupna je preko Interneta.

Osnova Blaise sistema je Blaise jezik koji se koristi za specifikaciju strukture i sadržaja upitnika. Blaise upitnik sastoji se obavezno iz delova *Quest* i *Route*, dok opciono može da sadrži delove *Check*, *Signal*, *Type*, *Variable* i *External*. Za složene upitnike mogu se definisati i delovi *Block* i *Table*. U sekciji *Quest* definišu se pitanja i skupovi mogućih odgovora. U sekciji *Route* definiše se staza (put ili skokovi) kroz upitnik, tj. određuju se pitanja koja treba da se postavljaju, po kom redosledu i pod kojim uslovima. U okviru sekcije *Check* definiše se kontrola (provera) konzistentnosti odgovora. Blaise programi mogu biti CADI za interaktivni unos i editovanje, CAPI za anketiranje uz pomoć *laptop* računara i CATI za telefonske ankete uz pomoć mikroracunara. CADI programom se nakon unosa upitnika aktivira opcija za kontrolu skokova i konzistentnosti. CATI i CAPI programi se tako pišu da se za vreme unosa kontrolišu skokovi i konzistentnost i da na ekranu monitora bude prikazano samo po jedno pitanje. U CATI i CAPI programima mogu da se definišu posebni blokovi (*Appointment block*) za pravljenje rasporeda anketiranja i vođenje evidencije o neodgovorima. U zavisnosti od složenosti upitnika program može imati više delova *Quest*, *Check* i *Route*, a podatke da smešta u jednu ili više Blaise datoteka (sekvencijalne indeksirane datoteke). Ove datoteke se mogu konvertovati u ASCII datoteke.

Deo Blaise sistema su MS-DOS programi za rukovanje podacima (Manipula), program za ponderisanje podataka (Bascula) i program za tabeliranje (Abacus).

U Institutu za razvoj resursa (IRD), sa prvenstvenom namenom obrade demografskih i zdravstvenih anketa na mikroracunarima, razvijen je Integrisan sistem za analizu podataka (ISSA - *Integrated System for Survey*

Analysis). U daljem tekstu detaljnije će biti opisan sistem ISSA, kao sredstvo novog načina obrade.

Nov pristup u obradi podataka demografskih i zdravstvenih anketa

Tokom 80-tih godina Institut za razvoj resursa (IRD) je bio angažovan na Programu demografskih i zdravstvenih anketa velikih razmera (DHS) u nerazvijenom delu sveta. Pri opredeljivanju za koncept obrade odlučeno je da se ona obavi na mikroračunarima. U tu svrhu je razvijen sistem ISSA.

Institut je uspešno, efikasno i kvalitetno, sproveo anketu u preko 35 zemalja. Uspehu su doprineli veoma dobar plan svih faza rada na anketi, kontrola njegovog sprovođenja i nov pristup u obradi. Nov pristup u obradi omogućen je prebacivanjem obrade iz velikih računskih centara na mikroračunare anketnih službi. Odlike ovakvog pristupa su: razvoj samošifrirajućih upitnika, unos podataka paralelno sa terenskim radom, inteligentan unos uz kontrolu struktura, raspona i konzistentnosti i obrada po manjim grupama upitnika (npr. po primarnim jedinicama uzorka).

Nov način obrade, sa unosom podataka za vreme terenskog rada, omogućio je blagovremeno otkrivanje grešaka u radu anketara i kontrolora i njihovo otklanjanje. Inteligentni unos podataka i obrada po manjim grupama, naročito su pospešili sekundarnu kontrolu podataka. Mnoge greške kucanja i šifriranja otkrivaju se i otklanjaju za vreme unosa tako da su liste grešaka u prolazima *batch* kontrole znatno kraće.

Nov način obrade IRD je sproveo u okruženju programskog paketa ISSA, koji je pomogao da se sve faze obrade, od unosa do pripreme izveštaja i tabela, strogo isprogramiraju i izvrše.

ISSA - obrada podataka na mikroračunarima

ISSA je moćan softverski paket za mikroračunare koji može da izvrši gotovo sve poslove obrade popisa ili anketa. On obezbeđuje unos, verifikaciju, logičku kontrolu i šifriranje podataka kao i kreiranje novih datoteka i tabela.

Startovanjem programa ISSA-e na ekranu monitora pojavljuje se glavni meni ISSA-e sa tri prozora sa nazivima: Rečnik (*Dictionary*), Aktivni rečnici (*Active dictionaries*) i Aplikacija (*Application*). U zavisnosti od prozora u kome se nalazimo, uz pomoć menija, funkcijskim tasterima moguće je odabrati: definisanje rečnika, editovanje aplikacija, izvršavanje aplikacija za unos ili *batch* obradu ili upotrebu nekog od uslužnih programa.

Centralna komponenta ISSA-e je *rečnik* koji služi za opis podataka ASCII datoteke. Za svaku datoteku koju obrađujemo, bilo da je kreiramo unosom, ili da ona već postoji, potrebno je da se definiše rečnik koji je u potpunosti opisuje. ISSA rečnici podsećaju na jezičke rečnike. Konstrukcija i upotreba su vođeni menijima. Svaki ISSA rečnik sadrži informacije o tipu datoteke (nivoima rečnika), njenoj strukturi (sekcijama rečnika) i pojedinim poljima datoteke (varijablama rečnika). Rečnik može da ima forme, nalik na forme upitnika, sa poljima za unos vrednosti sa upitnika.

ISSA je jedinstven po tome što može da obrađuje ASCII datoteke do na četiri nivoa. Tokom obrade glavne datoteke može da se pristupi eksternim datotekama sa kojih se, prema ključu, podaci mogu učitati u glavnu datoteku ili da se iz glavne u njih upišu. Dovoljno je, za sve faze obrade, jednom definisati rečnike podataka.

Za poslove obrade definišu se *aplikacije* koje upravljaju obradom delujući na podatke opisane rečnikom. U okviru aplikacija definišu se procedure pisane ISSA komadnim jezikom, korisničke datoteke sa porukama o

greškama i radni rečnik. Unutar ISSA sistema su procesori koji izvršavaju ISSA aplikacije.

Procedure aplikacije za interaktivni unos podataka, pozivajući se na odgovarajući rečnik, vrše kontrolu strukture, raspona, skokova i konzistentnosti po meri koja odgovara korisniku. Procesor za unos podataka izvršava procedure generišući ASCII datoteku podataka i indeksnu datoteku koja omogućava brz pristup unesenim slučajevima.

Batch procesor čita datoteku podataka sa ISSA rečnikom i obavlja sve akcije definsane procedurama za poslove obrade kao što su *batch* editovanje podataka, tabeliranje, kreiranje novih datoteka i pripreme izveštaja.

Logička i aritmetička konzistentnost može se proveriti ISSA komandnim jezikom. Poruke o greškama smeštaju se u izveštajnu datoteku. Korekcije se mogu obaviti interaktivno, od slučaja do slučaja, korišćenjem opcije ažuriranja procesora za unos, ili sistematski pomoću *batch* procesora. Imputacije podataka se takođe mogu obaviti *batch* procesorom.

ISSA može da proizvede tabele frekvenccije, tabele sa dva i tri ukrštanja i tabele sredina prema jednoj ili dvema nezavisnim varijablama. ISSA komandni jezik ima i naredbe za matrice operacije.

Od postojećih datoteka, ISSA može da izdvoji podskupove prema zadatim uslovima ili da formira datoteke sa novim šiframa vrednosti varijabli (*recoded*). Takođe, hijerarhijske datoteke mogu da se konvertuju u "ravne" i da se eksportuju u softverske pakete (npr. SPSS) koji ne mogu da obrađuju kompleksne strukture podataka.

ISSA sistem raspolaže i uslužnim programima za: štampanje rečnika, sortiranje podataka, generisanje indeksa, sabijanje datoteka brisanjem markiranih podataka i poređenje datoteka.

Programski paket ISSA, verzija 2.28+, napravljen je krajem osamdesetih godina. Za rad ISSA programa dovoljan je IBM/PC kompatibilan računar, 640 Kb RAM memorije, Hercules ili VGA monitor, hard disk

i štampač. Od softvera je potreban DOS 3.0 ili viši i naravno ISSA sistemske diskete.

Grafička rešenja ISSA-e, verzije 2.28+, su skromna u poređenju sa rešenjima koje nam nude danas raspoloživi softverski paketi za mikroracunare. Međutim, ISSA je specijalizovan paket za obradu podataka i na tom terenu je izuzetno moćna, fleksibilna i efikasna.

Institut za razvoj resursa (IRD) je pre nekoliko godina promenio ime u kompaniju Macro Internacional. Izašle su nove verzije ISSA-e koje podržavaju savremene hardverske i softverske trendove. ISSA, verzija 4, koristi produženu (*extended*) memoriju, radi na 386 ili jačim mašinama. Za praktičan rad potrebno je bar 2 Mb memorije. ISSA 4 umesto 940 varijabli, koliko je mogla da obrađuje ISSA 2, u jednoj aplikaciji može da rukuje sa čak 1800 varijabli. Povećan je broj tabela koje ISSA istovremeno može da obradi, jedino ograničenje je raspoloživa memorija. Raspoloživa memorija je takođe jedino ograničenje za dimenzije tabela.

ISSA 4 ima poboljšane mogućnosti izračunavanja frekvencija. Dok su u ISSA-i 2 mogle da se računaju frekvencije jednocifrenih i dvocifrenih numeričkih varijabli, u ISSA-i 4 frekvencije se mogu da računaju za sve varijable. Poboljšan je i pojednostavljen proces tabeliranja i izgled tabela je lepši. Dodata je nova komanda CROSSTAB koja omogućava kreiranje tabela "u letu", dok pišemo naše aplikacije. Mogućnosti izvoza podataka u druge aplikacije (SPSS, SAS) su takođe proširene i poboljšane. Osim toga vođeno je računa da ISSA 4 bude što je moguće više kompatibilna sa ISSA-om 2.

Primena ISSA-e u obradi podataka

Sistem ISSA je korišćen u obradi podataka nekoliko velikih svetskih anketa. Njene mogućnosti su potvrđene u obradi demografskih i zdravstvenih anketa, DHS (Cushing 1987, 1989). Takođe, obrada

podataka Ankete o porodici i fertilitetu (FFS) u zemljama ECE regiona (Ujedinjene Nacije, 1992, 1993 i Kližing, 1992-1996) sprovodi se u okruženju sistema ISSA.

U pisanju programa za unos i *batch* obradu podataka FFS istraživanja, u okviru obuke za rad sa sistemom ISSA u Varšavi 1992, učestvovao je autor ovog rada. S obzirom da se anketom FFS prikupljaju podaci o raznim aspektima porodice i fertiliteta primenjeni upitnik je vrlo složen. Sadrži detaljne istorije (biografije) promena sastava domaćinstva, bračnih stanja, intimnih odnosa, zatim istorije živorođenja i ostalih trudnoća, školovanja i aktivnosti ispitanika i sadašnjeg partnera (ako postoji). Pored ovih istorija događaja upitnik sadrži pitanja koja se odnose na stavove ispitanika i partnera o fertilitetu i formiranju porodice.

U sistemu ISSA bilo je moguće strogo kontrolisati unos podataka po ovom složenom upitniku. Osim kontrole raspona i skokova program za unos ankete FFS obezbeđuje kontrolu konzistentnosti odgovora u upitniku. Posebno se kontroliše konzistentnost datuma događaja. Proverava se datum završetka školovanja i starosti, starost pri rođenju prvog deteta, razmak između porođaja, razmak između poslednjeg događaja i datuma intervjuja itd.

U *batch* obradi u sistemu ISSA izvršene su korekcije i kontrole koje zbog složenosti nisu sprovedene pri unosu. Kontrolisani su kontingenti i izvršena imputacija datuma za događaje za koje su nedostajali. Osim toga u ISSA-i su urađene tabele i izveštaji FFS istraživanja.

Iskustva autora u pisanju aplikacija za unos i *batch* obradu za Anketu o porodici i fertilitetu u Zaječaru (Avramov i sar, 1992; Melovski Trpinac, 1996), kao i u pripremi programa za unos mnogih kraćih anketa, ne obavezno demografskih, su pozitivna. U većini anketa ozbiljniji problemi su, pre svega, bili vezani za nedovoljno poklanjanje pažnje organizatora pripremanja istraživanja. Posebno, često sadržaj i forma upitnika nisu prilagođeni savremenom načinu obrade. Takođe, značaj unosa što "čistijih" podataka se nedovoljno sagledava, te se mnoge kontrole podataka bespotrebno ostavljaju tek za *batch* obradu.

ISSA I BLAISE - Paralela

Iako oba sistema omogućavaju obavljanje poslova vezanih za obradu anketnih podataka, razlozi njihovog razvoja su različiti. Blaise je razvijen sa ciljem kompjuterizacije i integracije statističkih procesa statističkih službi. Razvojem ISSA-e težilo se stvaranju okruženja u kome bi bila moguća kompletna obrada složenih anketa, pre svega demografskih i zdravstvenih. Otuda proističu i njihove osnovne razlike.

Osnovu ISSA sistema čini rečnik, tj. opis sloga podataka. ISSA prihvata vrlo složenu strukturu podataka, do četiri hijerarhijskih nivoa, i podelu podataka na sekcije prema tipu podataka ili pak tematskoj celini. Pozivajući se na odgovarajući rečnik, aplikacija za unos vrši kontrolu skokova i konzistenosti za upitnike koji se unose. Pri tome kreira samo jednu ASCII datoteku sa podacima. Za pitanja koja su u upitniku preskočena (jer se ne odnose na određenog ispitanika) odgovarajuća polja u datoteci će biti nepopunjena. Za sekcije koje se čitave preskaču, ne odvaja se poseban prostor u datoteci. Sadržaj datoteke sa podacima se, uz mapu rečnika, lako može pročitati.

Osnovu Blaise sistema čini Blaise upitnik sastavljen od delova *Quest*, *Route* i *Check*. Deo *Route* određuje koja pitanja, u kom redosledu i pod kojim uslovima se pitaju, kao i redosled po kome se podaci smeštaju u datoteku. Za složenije upitnike definišu se blokovi *Subfiles* čiji se podaci smeštaju u posebne datoteke, može ih biti do 50. Pri tome Blaise datoteke sa podacima nisu ASCII datoteke.

Blaise sistem razlikuje CATI, CADI i CAPI aplikacije, dok u ISSA-i takva podela aplikacija nema značaja. ISSA aplikacija za unos, uz ISSA komandni jezik, može da bude sa karakteristikama bilo CATI, CADI ili CAPI aplikacije.

ISSA *batch* aplikacije, omogućavaju da se podaci uneseni ISSA programom za unos, ili postojeće ASCII datoteke sa podacima, za koji

je prethodno definisan rečnik, dalje obrađuju. Pored ostalog, moguće je obaviti kontrolu kontingenata, logičku kontrolu, korekcije i imputacije podataka, izdvajanje podskupova i tabeliranje podataka. U ISSA aplikacijama za *batch* obradu koristi se isti komandni jezik kao u aplikacijama za unos, uz mogućnost korišćenja nekih specifičnih komandi vezanih za *batch* obradu (npr. komande vezane za izračunavanje frekvencija, tabeliranje podataka, operacije sa matricama). Naravno, komande specifične za unos se isključuju. Za pisanje aplikacija u ISSA-i potrebno je razumevanje razlika između interaktivnog i *batch* režima obrade.

Blaise ne obuhvata kompletan proces obrade. Na primer, Blaise nema mogućnost *batch* verifikacije, kontrole i imputacije podataka u onoj meri u kojoj je to moguće u ISSA-i. U okviru Blaise sistema postoje posebni programi za ponderisanje i tabeliranje. U ISSA-i su ponderisanje i tabeliranje samo deo poslova koji se mogu obaviti u *batch* obradi.

U ISSA-i postoji mogućnost definisanja vremenskog trenutka izvršenja naredbi vezanih za nivo, sekcije ili varijable rečnika. Nešto slično procedurama PreProc (pre obrade, pre unosa) ili PostProc (posle obrade, posle unosa) ne postoji u Blaise-u.

Zaključne napomene

Uspех obrade podataka složenih anketa, meren kvalitetom obrađenih podataka i brzinom finaliziranja obrade i objavljivanjem rezultata, zavisi od niza činilaca. Neki od njih su:

- Dobar plan svih poslova ankete i obrade i kontrola njihovog sprovođenja
- Saradnja lica uključenih u rad na anketi i zajedničko rešavanje problema koji se javljaju u obradi
- Primena samošifrirajućih upitnika

- Obrada na mikroracunarima sa unosom podataka za vreme rada na terenu
- Primena inteligentnog unosa
- Obrada po manjim grupama upitnika

Razvijeni su softverski sistemi specijalno namenjeni za obradu podataka na mikroracunarima, među kojima Blaise i ISSA zauzimaju istaknuto mesto. U okruženju integrisanog sistema ISSA moguće je obaviti sve poslove obrade kompleksnih demografskih i zdravstvenih anketa, od unosa, verifikacije i kontrole, do pisanja izveštaja i tabeliranja. Inteligentnim unosom u ISSA-i datoteka sa podacima je gotovo bez grešaka odmah nakon što je unos završen i može se koristiti za preliminarno tabeliranje i analizu, umesto da bude u stanju da se sve greške strukture, raspona i skokova tek treba da otkriju i isprave.

U perspektivi, tehnološka dostignuća će i dalje bitno uticati na proces prikupljanja i obrade podataka. Skupa i vremenski zahtevna anketna istraživanja sa upitnicima na listu papira zameniće istraživanja sa *laptop* računarima ili telefonske ankete uz pomoć računara. Brz razvoj elektronskih mreža pružiće nove mogućnosti za prikupljanje, obradu i razmenu podataka.

Literatura

AVRAMOV, D. i sar. (1992). *Upitnik za anketu o porodici i fertilitetu*. (Zaječar).

BAKER, R. P. i dr. (1995). "Computer - assisted Personal Interviewing: An Experimental Evaluation of Data Quality and Costs". *Journal of Official Statistics*, Vol 11, No.4, 393-414.

BETHLEHEM, J.G. i M.H. Schuerhoff (1990). *Blaise 2.3 Tutorial*. Netherlands Central Bureau of Statistics.

CUSHING, J. (1987). "Processing the Demographic and Health Surveys: The Micro Approach". Invited paper of the 46th Session of the ISI, Booklet Recueil, (Tokyo).

- CUSHING, J. (1989). *Demographic and Health Surveys Data Processing Manual*. Institute for Resource Development/Macro International, Inc. (Columbia, USA).
- CUSHING, J. (1992). *ISSA - The Integrated System for Survey Analysis, Version 2.28 (Manual)*. Institute for Resource Development/Macro International, Inc. (Columbia, USA).
- KLIJZING, E. (1992-1996). *FFS Newsletters*. Economic Commission for Europe. (Geneva: United Nations).
- MELOVSKI TRPINAC, O. (1996) Anketa o porodici i fertilitetu, nepublikovane tabele, Zavod za zaštitu zdravlja u Zaječaru.
- SCHUERHOFF, M. H. (1996). A new version of Blaise. Communications. *Netherlands Official Statistics*, Vol. 11, summer, 57-58.
- *** (1990). "Data Editing in Federal Statistical Agencies", *Statistical Policy Working Paper 18*. Subcommittee on Data Editing in Federal Statistical Agencies, Federal Committee on Statistical Methodology.
- UNITED NATIONS, (1992). *Questionnaire and codebook for Fertility and Family Surveys in Countries of the ECE Region*. (New York).
- UNITED NATIONS, (1993). *Standard Recode Files and Standard Country Reports for Fertility and Family Surveys in Countries of the ECE Region*. (Geneva).
- MACRO INTERNATIONAL Inc., (1996). *Using VIEW in ISSA 4*. (Calverton, Maryland).

Olga Melovski Trpinac