

nekateri vidiki uporabe računalnikov v sociologiji in politologiji

c. trampuš
a. ferligoj

UDK 681.3:(301 + 32)

Fakulteta za sociologijo,
politične vede in novinarstvo
Univerze v Ljubljani

Podan je kratek pregled problemov pri uporabi računalnikov v sociologiji in politologiji. Opisane so smeri razvoja bank podatkov in metod.

SOME ASPECTS OF COMPUTER APPLICATION IN SOCIAL SCIENCES. A short survey of the problems involved in the application of computers in social sciences is given and some perspectives in the concepts of data archives and methods are described.

Ena izmed karakterističnih posebnosti sedanje znanstveno-tehnične revolucije je rastoča uporaba matematičnih metod na različnih področjih znanosti in tehnike. Ta proces, ki je upravičeno dobil naziv matematizacija znanstveno tehničnega znanja, obsega sedaj tudi taka področja znanosti, za katera je veljalo, da zaradi težavnosti in kompleksnosti problematike, raziskave z matematičnimi metodami niso možne (biologija, lingvistika, ekonomija, ...). Dejansko pa težavnost proučevanja pojavov v takih znanostih ne le, da ne izključuje, temveč nasprotno, zahteva oziroma predpostavlja uporabo točnih in abstraktnih matematičnih metod (Ruzavin, 1977).

V zadnjem času je tudi v družboslovju vedno bolj čutili potrebo po kvantitativnem preverjanju predpostavk o družbenih zakonitostih. Še do nedavnega je bilo empirično raziskovanje v sociologiji in politologiji omejeno le na zelo preproste statistične obdelave na maloštevilnih podatkih, saj bi kompleksne analize večjega števila informacij zahtevale ogromno naporov in časa. Ugodnejše pogoje za kvantitativno proučevanje pojavov v družbenih znanostih nudi šele sodobna računalniška tehnologija, ki omogoča:

- a. sistematično urejanje in izmenjavo velikega števila informacij, ki se pojavljajo vzporedno z vsakdanjim delom ljudi in se lahko zbirajo več let ali celo desetletij;
- b. uporabo ustreznih računalniško orientiranih matematičnih metod.

Problemi pri uporabi računalnikov

Družbene znanosti se ukvarjajo z realnimi vsakdanjimi pojavi, ki jih je težko prevesti iz vsakdanjega nedoločenega jezika v natančni jezik znanosti. "Meritve", v kolikor so sploh mogoče, so obremenjene z vrsto slučajnosti in mnogimi napakami, ki jih je potrebno oceniti. To je zelo težka naloga, saj naletimo na vrsto omejitev pri izvedbi eksperimentov in jih v večini primerov ne moremo ponoviti pod enakimi pogoji. Poleg tega se družbeni pojavi spreminjajo s časom in so v svoji medsebojni povezanosti in raznovrstnosti manifestiranja "neskončni". Neposredno je pri raziskovanju določenih pojavov mogoče opazovati le majhne izseke, za kar pa je največkrat potrebno zbrati veliko število podatkov (od katerih se nekateri izkažejo kot nepomembni), da dobimo relevantne informacije v največji možni meri.

Obilje računalnikov, različnih matematičnih metod za statistično obdelavo podatkov in podatkov samih, lahko pripelje do formalne uporabe matematičnih metod in do situacij, ko raziskovalec izgubi iz vida osnovni namen numeričnih manipulacij. Obdelava empiričnega gradiva ne more nadomestiti globokega poznavanja raziskovanih družbenih pojavov. Predstavlja le dopolnilo k premišljeno postavljeni nalogi o proučevanju pojavi in izboru ustreznih metod. Žal so te teze v družboslovnih raziskavah včasih neupravičeno pozabljene (Ajvazjan, 1974).

vanje skupine sorodnih problemov iz določenega področja.

Za statistične analize v družboslovju je v svetu in pri nas že vrsta "programskih paketov" kot so na primer: SPSS, STATJOB, P-STAT, SS, OSIRIS itd. Večina naštetih programov omogoča raznovrstne transformacije in izbiro podatkov; pogojno, brez-pogojno in slučajno, izbiro enot in vrsto statističnih in drugih analiz, ki so natančnejše našteje v naslednjem poglavju. Natančen pregled drugih programov, primernih za družboslovne raziskave je podan v reviji: *Social Science Information* 13(1974)3 na straneh 105-146.

2. Metode in programi za obdelavo podatkov v družboslovju

Analiza podatkov brez uporabe razvitejših matematičnih metod, omejena le na tabele in grafične ponazoritve, največkrat ne izčrpa vse informacije v podatkih o proučevanih družbenih pojavih. Zato si vedno več družboslovcev prizadeva poiskati primernejše tehnike za kvantitativno analizo podatkov. Računalnik omogoča uporabo postopkov z zahtevnejšimi numeričnimi prijemi (določitev lastnih vrednosti, reševanje enačb, ...) pa tudi uporabo postopkov, ki so osnovani na heurističnih metodah, kjer je potrebno pregledati veliko možnih izidov.

Statistične metode, ki so primerne za analizo podatkov v družboslovju, je mogoče urediti v skupine podobnih metod. Tak izčrpen pregled metod primernih za analizo podatkov v političnih vedah je podal Alker (1975). Podobna klasifikacija je bila izdelana v letu 1976 na Mednarodnem delovnem seminarju SOECO v Ljubljani in na Simpoziju računalniške statistike COMPSTAT v Berlinu. Na grobo lahko metode empiričnega proučevanja v družboslovju razdelimo na dve skupini: metode za merljive spremenljivke in metode za kvalitativne značilnosti.

a. Metode za analizo merljivih spremenljivk

Statistična analiza ene ali dveh merljivih spremenljivk je poznana iz klasičnih učbenikov statistike (Blejec, 1973, Jamnik, 1976). Vse pogosteje se v družboslovju za preverjanje zapletenejših trditev o povezanostih večih merljivih spremenljivk uporabljajo multivariatne metode kot parcialna in multipla korelacija, multipla linearna regresija, analiza va-

riance, faktorska analiza, kanonska korelacijska analiza in diskriminacijska analiza (Anderson, 1958, Cooley in Lohnes, 1971, ...). Za omenjene analize imamo na voljo v Računskem centru na FSPN nekaj računalniških programov kot SPSS, SS, STATJOB, P-STAT, BMD, ... Testiranje razlik faktorskih struktur in podobnih zapletenejših hipotez (Jöreskog, 1970) omogočata programa SIFASP in ACOVS. Pomemben dosežek za empirično proučevanje v družboslovju so vzročni linearni modeli (Blalock, 1964, 1971, Golderger in Duncan, 1973, ...), s katerimi je mogoče preverjati zapletenejše vzročne zakonitosti v družboslovju. Tak, dokaj splošen vzročni model, ki vključuje kot posebne primere rekurzivni, nerekurzivni in faktorski model, je vgrajen v program LISREL.

b. Metode za analizo kvalitativnih lastnosti

Empirično proučevanje družboslovnih zakonitosti je za razliko od naravoslovnih težavnejše predvsem zato, ker so ponavadi proučevane lastnosti le kvalitativnega značaja (ordinalne ali nominalne lestvice) in zanje povečini še niso izdelani primerni matematični pristopi. Nekaj klasičnih neparametričnih testov za take lastnosti je vključenih v paketih SPSS, STATJOB in P-STAT. V zadnjem času pa je tudi na tem področju predlaganih nekaj pomembnih rešitev. Psihologi Torgerson (1958), Coombs (1964) in drugi so izdelali metode lestvičenja (SPSS, SCALE). Tem so sledile metode večrazsežnostnega lestvičenja (Torgerson, 1958, Shepard, 1962, Kruskal, 1964), ki iz danih podobnosti med enotami določijo lego teh enot v Evklidskem prostoru s čim manj razsežnostmi tako, da je napetost med njimi čim manjša. Za te metode sta dostopna programa MDSCAL in MINISSA. Za družboslovne raziskave je še posebej zanimiva splošnejša metoda večrazsežnostnega lestvičenja Carolla in Changa (1970), ki analizira spremembe v zaporedju matrik podobnosti (INDSCAL). Za proučevanje kvalitativnih značilnosti se vse bolj uveljavljajo tudi metode določanja skupin (Cluster metode). Programi zanje so ponavadi hitri in jih lahko uporabljamo za večje količine podatkov. Najpogosteje se uporabljajo postopki hierarhičnega določanja skupin, pri katerih je mogoče proces združevanja enot v skupine grafično predstaviti z drevesom - dendrogramom. Za te postopke so dostopni trije programi: HICLA, HICAJOHN in CLUSE. Slednji temelji na Lance-Williamsovem

Pregled in perspektive razvoja banke podatkov in metod

Izkušnje kažejo, da se razvija uporaba računalnikov v družbenih vedah med drugim tudi v naslednjih smereh:

1. Formiranje računalniško vodenih bank podatkov

Dosedanja praksa v svetu in posebno pri nas kaže, da so podatki iz večine raziskav neurejeni, neustrezno arhivirani in niso na razpolago vsem uporabnikom na tem področju. Ker ostajajo tako podatki kot rezultati (pri tem zlasti številni materiali, ki niso vključeni oz. upoštevani v tekstovnih interpretacijah) raztreseni med številnimi posamezniki, institucijami in disciplinami, prihaja do ponavljanja istovrstnih vprašanj v raznih anketah in do medsebojno nepovezanega proučevanja teh vprašanj, kar vodi do nesmiselnega trošenja energije raziskovalcev. Sistematizacija podatkov vodi k večji selektivnosti tako, da se jasneje nakazujejo predvsem tista vprašanja, ki so še nepojasnjena in tista področja, pri katerih bi ponovno zbiranje informacij pomenilo dodatno trošenje finančnih sredstev.

Urejeni arhivi naj bi napravili korak naprej k temu, da bi tudi empirične ugotovitve lahko privedle do pomembnejših teoretičnih posplošitev tj. do sinteze empiričnega in teoretičnega obravnavanja posameznih vprašanj. Sedanja razpršenost in izredno težka dostopnost rezultatov in podatkov iz že opravljenih raziskav obenem določa meje možnega teoretičnega posploševanja. Vpogled v rezultate dosedanjega dela je odvisen povsem od številnosti osebnih stikov in poznanstev, ne naslanja pa se na kakršen koli sistematično organiziran in preverjen fond podatkov. Posamezen raziskovalec mora torej nesorazmerno veliko količino časa posvetiti povsem slučajnostnemu neorganiziranemu iskanju in preverjanju relevantnih virov.

Poseben vidik, ki zaostruje potrebo po standardizaciji podatkov in njihovem sistemiziranju predstavlja proučevanje dinamičnih sistemov, ugotavljanje različnih trendov oz. zakonitosti glede povezanosti med pojavi oz. procesi v času in prostoru. Dinamičnih sistemov ni mogoče obravnavati brez kontinuitete virov informacij oz. načinov konceptualizacije v posameznih časovnih točkah. Tudi s tega vidika in zlasti z vidika marksističnega metodološkega pris-

topa se pojavlja upravičena kritika sedanje raziskovalne prakse, ki se predvsem omejuje na statične preseke družbenega dogajanja in s tem vodi do enostranskih interpretacij.

Sistematično urejeni in dosegljivi podatki predstavljajo skupaj z računalniško vodeno dokumentacijo tudi študijsko gradivo pri pouku sociologije in političnih ved ter informacijsko bazo za druge uporabnike.

Nadaljnja naloga arhivov podatkov je standardizacija informacij na področju sociologije, politologije in komunikologije. Za razliko od sedanjega stanja, ko vsak raziskovalec po svoje opredeljuje pojme in spremenljivke, ki jih uporablja, naj bi v prihodnje poiskovali vsaj do neke mere standardizirati tiste spremenljivke, ki se najpogosteje pojavljajo v empirično raziskovalnem delu. Le v tej meri lahko iščemo možnosti zagotavljanja širših teoretičnih posploševanj na osnovi pri nas ugotovljenih empiričnih podatkov. Urejeni arhivi bodo prispevali k zmanjšanju količine potrebnih informacij tudi glede na to, ker bodo evidentirali uporabljena kompleksna in sintetična merila posameznih sklopov medsebojno povezanih pojavov in pomembnost posameznih indikatorjev oz. spremenljivk. Selektivnost v tej smeri bo torej pomenila, da bi z manjšim številom podatkov dosegli enako ali celo pomembnejše znanje in lažjo možnost obravnave dinamičnih sistemov oz. predvidevanja sprememb v prihodnosti. Nadaljnji razvoj naj bi šel v smeri tesnejšega povezovanja številnih podatkov iz anketnih raziskav (predvsem informacij o različnih percepcijah in dispozicijah posameznikov) z obsežnim gradivom, ki ga nudi bodisi redna statistična služba v posameznih sektorjih, bodisi občasne statistične akcije in drugi arhivi pri nas in v inozemstvu.

V razširjeni koncept arhivov podatkov spadajo tudi priprave in razvoj sistemov za shranjevanje in izkazovanje podatkov. Posebno važna nadaljna naloga arhivov je zbiranje in razvoj programov za računalnike, tako za manipulacije s podatki, kot tudi za reševanje statističnih in drugih obdelav podatkov. Programi morajo biti pripravljeni tako, da so na najenostavnejši način dostopni vsem interesentom kot so npr. učitelji in raziskovalci, študentje in drugi. V zadnjem času teži razvoj k izdelavi in izpopolnjevanju že obstoječih programov, ki omogočajo reše-

(1967) obrazcu in vsebuje zmožnosti prvih dveh programov (Johnson, 1967) kot poseben primer. Obeta-joči so postopki, ki temeljijo na razpoznavanju vzorcev (Žuravljov, Kamilov in Tuljaganov, 1974), za katere je v pripravi precej splošen računalniški program, katerega del je dostopen tudi pri nas (PARE). Ti postopki rešujejo za kvalitativne podatke podobne probleme kot znani multivariatni postopki za merljive spremenljivke.

Z obema skupinama metod se delno prepletajo še druge metode, kot na primer slučajni procesi (Coleman, 1964, Kemeney in Snell, 1962, Bartholomew, 1967, Fararo, 1973), simulacije (Dutton in Starbuck, 1971), ...

L i t e r a t u r a :

- Ajvazjan C.A., Bežajeva Z.I., Staroverov O.V.: Klasifikacija mnogomernih nabljudenij, Statistika, Moskva 1974
- Alker Jr. H.R.: An overview of polimetrics, Handbook of political science, Volume 7, 1975
- Anderson T.W.: An Introduction to Multivariate Statistical Analysis, John Wiley, New York, 1958
- Blalock H.M.: Causal Inferences in Non-experimental Research, Chapel Hill: University of North Carolina Press, 1964
- Blalock, H.M.: Causal Models in the Social Sciences, Aldine Publishing Company, Chicago, 1971
- Blejec M.: Statistične metode za ekonomiste, Univerza v Ljubljani, Ljubljana, 1973
- Caroll, J.D., Chang, J.J.: Analysis of individual differences in multidimensional scaling via a N-way generalization of "Eckart-Young" decomposition, Psychometrika, 1970
- Coleman, J.S.: An Introduction to Mathematical Sociology, The Free Press, New York, 1964
- Cooley W.W., Lohnes P.R.: Multivariate Data Analysis, John Wiley, New York, 1971
- COMPSTAT 1976 Proceedings in Computational Statistics, Physica-Verlag, Wien 1976
- Coombs, C.H.: A Theory of Data, John Wiley, New York, 1964
- Dutton, J.M., Starbuck, W.H.: Computer simulation of human behaviour, New York, John Wiley, 1971
- Fararo, T.J.: Mathematical Sociology, John Wiley, New York, 1973
- Goldberger, A.S., Duncan, O.D.: Structural Equation Models in the Social Sciences, Seminar Press, New York, 1973
- Jamnik R.: Uvod v matematično statistiko, DMFA, Ljubljana, 1976
- Johnson, S.C.: Hierarchical Clustering Schemes, Psychometrika, 1967
- Lance, G.N., Williams, W.T.: A general theory of classificatory sorting strategies, 1. Hierarchical systems, The Computer Journal, 9 (1967) 4, 373 - 380
- Kemeney, J.G., Snell, J.L.: Mathematical Models in the Social Sciences, Ginn, Boston, 1962
- Kruskal, J.B.: Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis, Psychometrika (I in II), 1964
- Mlinar Z.: Dialektika, empirizem in spekulacija v socioloških in politoloških raziskavah. Referat na posvetovanju: "Teoretično-metodološke osnove raziskovanja v družbenih vedah", ki ga je organiziral Marksistični center CK ZK Srbije v Beogradu, decembra 1976
- Ruzavin, G.I.: Matematizacija naučnega znanja, "Znanie" (Serijs "Filozofija"), Moskva 1977
- Shepard, R.N.: The Analysis of proximities: Multidimensional scaling with an unknown distance function, Psychometrika (I in II), 1962
- SOECO, Report of the methodological Group, FSPN, Ljubljana 1976
- Torgerson, W.S.: Theory and Methods of Scaling, John Wiley, New York, 1958
- Toš N.: Metode družboslovnega raziskovanja, FSPN, 1975
- Žuravljov, Ju.I., Kamilov, M.M., Tuljaganov, Š.E.: Algoritmi vičisenija ocenok i ih primenenie, "Fan" Taškent, 1974
- Bartholomew, D.J.: Stochastic Models for Social Processes, John Wiley, New York, 1967
- Jöreskog, K.G.: Simultaneous Factor Analysis in Several Populations, Research Bulletin 70-61, Princeton, N.J.: Educational Testing Service, 1970