

Elmélet és módszertan

Történeti kutatómunka – gombnyomásra? *Beszélgetés a SZOCPROG rendszerről*

Granasztói György: Az elmúlt években nagy sikerrel működött a Kvantum Klub az Akadémia várbeli központjában. Számítógépet használtak kutatók – akik zömmel az Akadémia társadalomtudományi kutatóhelyein dolgoznak – találkoztak itt rendszeresen, hogy munkabeszámolókat tartsanak és véleményt cseréljenek. A „Kvantumban” ismerkedtünk meg többek között a SZOCPROG társadalomtudományi programcsomaggal, amely lényegében egy adatelemzési rendszer, s amely a közelmúltban a szélesebb nyilvánosság elé is kilépett. Beszélgetésünket részben ez az esemény tette időszerűvé. A SZOCPROG kidolgozóit, Manchin Róbert és Füstös László egyhetes intenzív tanfolyamot szerveztek a Várban, több mint hatvan érdeklődő számára. Megismerkedhettünk a programmal, megtanulhattuk a használatát, de ami ennél is fontosabb, képet kaptunk arról, milyen azonnal használható kvantifikációs technikák állnak ma itthon rendelkezésre, ha különböző tömeges statisztikai adatokat tartalmazó forrásokat kívánunk elemezni. A Manchinnal és Füstössel folytatandó beszélgetésben azt kíséreljük a Történelmi Szemle olvasói elé tárni, hogy mire alkalmas a SZOCPROG, ezt követően a történeti adatelemzésben jártas Szabó Dániellel és Göndör Györggyel folytatjuk az eszmecserét.

Előjáróban még el kell mondani, hogy a tanfolyam gazdag programot kínált. Az elhangzott előadások között szerepeltek például a következő témák: Statisztika és számítástechnika a társadalomtudományokban; Statisztikai alapfogalmak; Számítástechnikai alapismeretek a CDC 3300-as gépen; Adatelőkészítés, adattranszformáció, adatstruktúrák, adatmanipulációk lehetősége; Empirikus elemzések, Regresszió elemzés, Korreláció elemzés, Faktor elemzés, Klaszter elemzések és így tovább. Vagyis a programrendszer ismertetése mellett az ebben a rendszerben szereplő különböző, döntően már a mérésel és a számítások matematikai statisztikai megvalósításával foglalkozó elméleti problémák is szerepeltek.

Az első kérdés, amelyet fel szeretnék tenni, a következő: Milyen kutatói igényt láttak maguk előtt a SZOCPROG-rendszer megalkotásakor?

Manchin Róbert: Mielőtt rátérek a válaszra, szeretném vázolni, milyen problémákat láttunk a számítástechnikának a társadalomtudományokban való felhasználásával kapcsolatban. Egyfelől a hetvenes évek elején kiderült, hogy különösen a szociológiában, de más tudományágakban is tömeges adatfeldolgozást nem tudnak számítógép nélkül elvégezni. A gyakorlat akkor általában az volt, hogy hamarabb tudták „kézzel” megoldani a feladatokat, mint ameddig végigszenvedték azt az elemzési fázist, amit egy új számítógép, egy új programozóval való közös nyelvkeresés, megértés, megértetés jelentett, és ezért az első próbálkozások lelkesedése után kijózanodás következett. Az eredmény az lett, hogy az első kezdeményezések tulajdonképpen le is álltak. Másfelől az érintett kutatógárda képzéséből még teljesen kimaradt mindenfajta számítástechnikai kérdés, így az érzékenységük is csekély volt az ilyen típusú ismeretek iránt. Azt hiszem, még most is az a helyzet, hogy bár formálisan szerepel az egyetemi tanrendben a számítástechnikai ismeret, és így talán a történések is, de például a pszichológusok bizonyosan, két félévet hallgatnak programozás és kibernetika címén, ám mindamellert az egyetemeknek nincsen sem anyagi eszközük, sem hozzáférésük a számítógéphez. A végzős hallgatóknak nincs ilyen természetű gyakorlata, tudása teljesen meddő. Ennek ellenére a fiatalabb generáció jó ideje kereste a módot, hogyan lehetne mégis korszerűen feldolgozni és elemezni az összegyűjtött adatokat. A nemzetközi szakirodalomból is kiolvasható volt, hogy az újabb elemzési technikák alkalmazása az elméletépítésben vagy az adatokból nyerhető összesítések terén mennyi mindent kínál. Amikor Szelényi Ivánnal elkezdtünk hármásban dolgozni, nagy vizsgálatba fogtunk, és hamar rájöt-

tünk, hogy egyszerűen nem tudunk dolgozni a hagyományos eszközökkel, munkánkhoz új technikai hátteret kell teremtenünk. Két út volt lehetséges. Az egyik, hogy átveszünk egy létező, működőképes rendszert. Két ilyen jött számításba. Ezek valóban elterjedtek, hajlékonyak és jól is használhatók. Az egyik rendszer az SPSS, a másik az OSIRIS. Mind a kettőt alkalmazzák, nemcsak szociológusok, hanem közgazdászok, történészek, mások is. Amikor elkezdtünk tájékozódni, kiderült, hogy az OSIRIS-t eleve IBM gépekre dolgozták ki, az akadémiai CDC 3300-as gépre való átalakítása igen nagy munka lett volna. Az SPSS az OSIRIS-nél valamivel egyszerűbb, szintén szélesben elterjedt, ám a memóriaigény akkora, hogy – bár létezik CDC változata is – a számunkra akkor lehetséges CDC konfiguráció mellett nem lehetett alkalmazni. (Időközben a SZTAKI megvette az SPSS rendszert. – A szerk.) Ekkor kezdtünk tulajdonképpen a másik út követésén gondolkodni, azon, hogy a saját munkánknál alkalmazandó különleges rendszert lehetőleg minden hasonló elemzésnél újra fel lehessen használni, továbbá, hogy a rendszer kezelése legyen minél egyszerűbb. Sikertült fellelkesíteni egy SZTAKI-s kollegát, Tóth Károlyt, akinek a SZOCPROG rendszer nagyon sokat köszönhet. Elmondtuk neki, hogy tapasztalatunk szerint milyennek képzeljük azt a kutatót, aki szeretné megérteni magát az elemzési módszert, szeretné alkalmazni is, de úgy érzi, nincs ereje arra, hogy megpróbálja elsajátítani az egész szükséges számítástudományi apparátust is. A célunk tehát az volt, hogy olyan természetes nyelvű paramétereket állítsunk össze, amelyekkel egyszerűen lehet megfogalmazni elemzési kéréseinket. Például ha korrelációt akarunk számítani az iskolai végzettség és a jövedelem között, akkor ezt a kérést úgy fogalmazzuk meg a gép számára, hogy leírjuk: „korreláció”, „iskolai végzettség”, „jövedelem”. Az eredmény az lesz, hogy bizonyos fix információk birtokában, de tulajdonképpen egyedül e szavak közlésével a kérdező megkapja a kívánt elemzési eredményt.

Füstös László: Azt vártuk a program-rendszertől, hogy a számítógép segítse végrehajtani azt a logikai sort, amit a kutató adatelemzései során technológiaként előír magának. Legyen képes a rendszer a modellezésnek azokat a szakaszait végigvinni, amit idáig a kutatók egyenként végeztek. A modellalkotás folyamatát, ahogy Manchin említette, azokkal az egyszerű kulcsszavakkal lehessen irányítani, amelyeket a kutató használ elemzéseiben. Tehát a rendszer maradjon a kutató nyelvi ismeretének szintjén, ne kelljen elsajátítania azt a programozói gyakorlatban használatos zsargont, a számítástechnika sajátos terminológiáját, amely miatt sok esetben nehéz kommunikálni a számítástechnikai szakemberrel. A célunk az volt, hogy a programrendszer értse meg a társadalomtudományi elemzéssel foglalkozó kutató kulcsszavait, működjön is ezek ismeretében, s a kutató feladata a kulcsszavak, tehát a modell elemeinek pontosabbá tétele és a megértése. A számítógép, mint eszköz, álljon a kutató szolgálatába.

Granasztói: Szeretnék egy közbevető kérdést feltenni, abból a célból, hogy világosan lássuk, miként is működik a rendszer. Tegyük fel – valószínűleg nincs így –, hogy Celldömölk hatósága a 19. század végén olyan felmérést készített, amelyben összeírták, egy-egy ház hány helyiségből áll, milyen mellékhelyiségei vannak, milyen gazdasági épületek tartoznak egy házhoz, mennyi a jövedelme a háztulajdonosnak, hányadmagával lakik ott és így tovább. Mit tegyen az az „érdeklődő” történész, aki a SZOCPROG segítségével szeretné elemezni a forrást?

Füstös: Miután föllelte ezt az adatforrást, és kiválasztotta az adatokat a halmazból – például egy úgynevezett adatlapra egyszerűen sorba írta őket – a számítóközpontban ezeket valamilyen adathordozóra (lyukkártyára, mágnesszalag, mágneslemez) viszik. Ezek az alapinformációk. Ha mondjuk nem önmagában a szobák száma érdekel, hanem az egy szobára eső személyek száma, akkor új mutatót kell képezni, és a programrendszernek azt kell mondani, hogy osszuk el a család létszámát a szobák számával. Ez a SZOCPROG nyelvén egyszerűen azt jelenti, hogy egy „GENERÁLÁS” szót írunk vagy lyukasztok le, melléírom az új változó nevét és képzési módját. A példában az egy szobára eső fők száma – mondjuk – a „laksűrűség” névre hallgat, amely tehát egyenlő a család létszáma osztva a szobák számával. Ha a továbbiakban azt vizsgálánk, hogy a jövedelem nagysága miként függ össze a laksűrűséggel, akkor a következő elem az lenne, hogy „KORRELÁCIÓ”. A feladat a SZOCPROG nyelvén is ugyanígy hangzik, vagyis ugyanezzel a kulcsszóval működtethető a program.

Granasztói: Visszatérve a kiinduló példára; amikor a kutató elkezdte ezt a munkát, tehát adatainak az elemzését, tulajdonképpen hány szót, milyen sorrendben kell felírnia a vezérlőkártyákra?

Füstös: A fenti példa a SZOCPROG nyelvén úgy oldható meg, hogy egy „ADAT” névre hallgató utasítással beolvassa az adatokat, ahol opcióként megmondhatjuk, szalagon, vagy mágneslemezen van-e az adat, ugyanezen az utasítás-kártyán elnevezhetjük a változóinkat, tehát például az első

adat a családnagyság, a második a szobák száma, a harmadik a jövedelem, a negyedik a sertések száma és így tovább. Ebben az utasításban felsoroljuk azokat a változókat (azokat az adatokat), amelyekkel dolgozni akarunk. A második utasítás például a „LEÍRÁS” utasítás lesz. Ennek hatására a felsorolt adatoknak az alapstatisztikáit fogja a rendszer elkészíteni, vagyis megkapjuk, hogy átlagosan milyen nagyságú családoknál, milyen nagyságú az átlagjövedelem, és ezeknek az adatoknak mekkora a szórása a mintánkban, mi a minimális, a maximális értéke. Ha korrelációkat akarunk számítani a fenti változók között, akkor az utasítás úgy szól: „KORRELÁCIÓ”, például felsoroljuk a változóneveket, és akkor a program kiszámítja a korrelációkat a rendszerben korábban definiált összes névre. Az „ÖSSZEZÉS” című utasítás eredménytáblázata alapján kaphatjuk például a különböző tagszámú családok átlagos jövedelmét. Más példa: a klaszterelemzésnél a „KLASZTER” szót kell írni utasításként és utána megmondani, hogy milyen eljárással kérjük a klaszterezést. A SZOCPROG jelenleg 18 különböző klaszterelemző eljárással rendelkezik.

Granasztói: Látjuk, egész kutatási stratégia jellemzi a SZOCPROG társadalomtudományi programcsomagot. A Füstös és Manchin folytatott beszélgetés után a kérdés az, tudja-e használni a történést a SZOCPROG-rendszert? Azt hiszem, jó néhány olyan esetet el lehet képzelni, amikor kielégítő, vagy legalább kezdeti eredményeket ad. Rá kell azonban mutatni arra is, hogy sok hasonló adatkezelő- és -elemző-rendszer létezik, kettőt Manchin Róbert is említett. Megemlíthetem még a francia FORCOD rendszert, amelyet történészek dolgoztak ki kifejezetten történeti problémák megoldására, és amelyet egyébként a Szemlében régebben már ismertettünk is. Mostani beszélgetésünkben a SZOCPROG rendszert példának tekintjük, hivatkozási alapnak, mert nem vállaltuk azt a feladatot, hogy előnyeit és hátrányait taglaljuk. Ha azonban tágabb összefüggéseket keresünk, vagyis arról beszélgetünk, miként befolyásolják a történeti munkát az új informatikai eszközök és módszerek, akkor azt hiszem, rá kell mutatni néhány általánosabb problémára. Az első általános probléma a következő: évszázadok óta, tulajdonképpen Bacon óta, a tudományos kutatásnak az a taktikája, hogy kérdésszerű kérdésre halad, lépésről lépésre próbálja megoldani a feladatát. Ez olyan taktikai elképzelés, amely a természettudományokra éppúgy jellemző, mint a társadalomtudományokra. Mintegy negyven-ötven éve vetődött fel, először a természettudományokban, de most már a társadalomtudományokban is, hogy ahol lehetséges, célszerű egyidejűleg több kérdésre keresni a választ, tehát olyan kutatási eljárást alkalmazni, amellyel egyszerre minden irányban lehet „dolgozni”. A számítástechnika elterjedése ezt az elképzelést viszonylag könnyen megvalósíthatóvá teszi, maga a SZOCPROG is képes e taktikát megvalósítani. A másik fontos szempont az, hogy az adat fogalma is megváltozik a számítógép megjelenésével és elterjedésével. A számítógép egy-egy formát hallatlanul rövid idő alatt, akár milliárdnyi elemi művelettel képes alávetni. Ha mondjuk egy régészeti intézet raktárában több száz vagy több ezer hasonló tárgy, például vázák, lapulnak, akkor ezekről a vázákról, éppen a számítógépes adatkezelés révén, olyan részletfinomságig lehet adatfelvételt készíteni, amelyre korábban egyszerűen nem volt mód. Itt a kutató helyzete is megváltozik, mert miközben a kísérleti tapasztalatok szintézisét végzi, igen gyorsan kezelhet egy mégoly nagy adattáblázatot. A kísérleti tapasztalatok lassú szintézise mindig a fejünkben fog lezajlani, s ezt a szintézist az emlékezet készíti elő. Mostantól azonban a fej, az emlékezet óriási segítséget kap egy lényeges szakaszon, ti. akkor, amikor szembesíteni kell az egyberakott adatokat. Ebben a helyzetben nem az a fontos, hogy kevesebbet kell dolgozni, hanem az, hogy olyan dolgokat is mérlegelünk, amelyeket azelőtt nem voltunk képesek „észben tartani”.

Még valamit szeretnék megjegyezni. Mindaz, amiről itt szó esik, nem emberfeletti. Nem egy monstrumról beszélünk, amely ráng fog telepedni, vagy felfal. Am éppen azért, mert emberi alkotás, nem árt tisztázni, kik hozzák létre. Szinte biztos, hogy a különböző történeti számítógépes alkalmazások, akarva-akaratlanul, más tudományok, elsősorban természettudományok közvetett hatása alatt keletkeznek. El kell gondolkodni azon, hogy ez a tény mennyiben befolyásolja a történeti megközelítés függetlenségét.

Szabó Dániel: Számomra, aki egyrészt a matematikai statisztika módszereiről, másrészt a számítógépről mind ez ideig csak homályos fogalmakkal rendelkeztam, a SZOCPROG tanfolyam mindenekelőtt tisztázta az ezekkel kapcsolatos fogalmakat.

A statisztikai módszereknek a konkrét programcsomag használatával összekapcsolt ismertetése nem keltette bennem azt az érzést, hogy már én is képes leszek tüstént, akár a jövő héten ilyen módszerekkel és eszközökkel dolgozni, azt ellenben nyilvánvalóvá tette, hogy a módszerekhez és az

eszközökhöz való intellektuális hozzáférés egyszerűbb, mint korábban véltem; érdemes tehát azon gondolkodni, tudom-e hasznosítani mindezt a kutatómunkámban. Vannak-e olyan kérdéseim, melyeket könnyebben tudok segítségükkel megoldani? Odáig azonban a tanfolyam végére sem jutottam, hogy feltételezzem, már vannak olyan kérdéseim, melyekre csak a matematikai statisztika és a számítógép segítségével kaphatok választ. (Más kérdés, mi a szerepe a számítógépnek, mint tömegforrás nyilvántartónak, mint pusztán technikai segédeszköznek a hagyományos kérdések vizsgálatánál.) Másrészt felmerült, hogy az új módszerek ismerete nem kelt-e bennünk új kérdéseket.

Tehát a tanfolyamon egyrészt a statisztikai módszerek ismertetését és konkrét alkalmazását tartottam fontosnak, másrészt a konzervcsomag olyan szempontú bemutatását, hogy egy sor adattípusra s egy sor kérdésre már minden különösebb és főleg hosszadalmasabb előkészület nélkül lehet kész programot kapni.

Mindez nyilvánvaló azok számára, akik eddig is dolgoztak statisztikai módszerekkel és használtak számítógépet. Sokan vannak a szakmában, akik ezt olvasmányaik stb. alapján elvileg tudomásul vették. A tanfolyam előnye mindezekhez képest az „érzéki megismerés” volt. Az a lehetőség, hogy majdnem kutatói szituációban találkozhattunk a jelzett dolgokkal.

Az ilyen tanfolyamok azonban a társadalomtudományokba egyre inkább benyomuló új módszerek és technikák egyik arculatát ismertetik csak meg. Kétségtelen, nagyon fontosak, s nem is pusztán azok számára, akik maguk is alkalmazni akarják mindezeket, hanem azok számára is, akik az új módszerek, technikák segítségével, felhasználásával készített munkákat olvasni fogják. Amiből az is következik, hogy a tájékoztatás szintjén, már az egyetemen tanítani kellene a történész hallgatóknak matematikai statisztikát, s némi elvi számítástechnikát is, mert különben továbbra is az lesz a helyzet, mint jelenleg: ha ilyen módszerrel készült munkákkal találkoznak, többnyire a jól bevált „ugorgyunk” technikához fognak folyamodni.

Bár a tanfolyam egyik fő célja az volt, hogy segítse mindazokat, akik matematikus és programozó közbeiktatása nélkül próbálják e módszereket alkalmazni, a SZOCPROG tapasztalatai birtokában is az a meggyőződés, hogy a megfelelő szakemberrel való együttműködés, konzultáció elengedhetetlen feltétele e módszerek és technikák optimális társadalomtudományi alkalmazásának.

Göndör György: Bizonyára vannak olyan társadalomtudományi feladatok, amelyek esetében a számítógépek alkalmazása megoldást jelent, míg másoknál nem. Ha valaki számítógépet alkalmaz, az még nem jelenti, hogy számítást végez, matematikai módszert használ. Hogy csak néhány példát említsék: egyszerű adattárolás, adatfeldolgozás, tágabb értelemben, szimuláció, folyamatok vezérlése stb. is lehetséges. Ha tehát a matematikai módszerek alkalmazását tekintjük, ez már a számítógép felhasználási területeinek szűkítését jelenti. További szűkítést jelent, ha csak a valószínűségi modellekre gondolunk. Léteznek számítógépes, nem valószínűségi matematikai alkalmazások is. Például közgazdasági feladatoknál gyakran olyan egyenleteket, egyenletrendszereket kell megoldani, amelyekben nincs valószínűségszámítás. Más társadalmi folyamatot is modellezhetünk, például differenciál egyenletekkel (hasonlóan természeti, műszaki folyamatokhoz). A valószínűségszámításnak a matematikai statisztika csak egy fejezete, ezen belül szűk részt alkotnak a hagyományos matematikai statisztikai módszerek, mint pl. a korreláció, a regresszió, a faktor analízis, a variancia analízis, a klaszter analízis. Ezek lényegében kész kidolgozott módszerek, amelyek használatáról elég nagy tapasztalat gyűlt össze. Mindegyiknek nagy irodalma van, lehet tudni, hol lehet ezeket a módszereket alkalmazni, hol nem (bár folynak kutatások abban az irányban, hogy új területeken hogyan lehet, ill. szabad-e ezeket alkalmazni). Ha tehát valaki, mondjuk egy történész, már átfogalmazta a problémáját matematikai statisztikai problémává, nem biztos, hogy talál majd kész módszert a megoldáshoz, sőt majdnem biztos, hogy nem. Ebből a helyzetből két irányba léphet tovább: vagy az eredeti kérdésközpontot változtatja, alakítja addig, amíg az már egy kész módszerrel megoldható. Magyarul a módszerhez keresi a feladatot. Vagy pedig kitart az eredeti problémafeltevés mellett, és új módszert próbál kidolgozni. Természetesen nem a – teszem azt – történésznek kell a felvetődött matematikai problémát megoldania, hanem a matematikusnak. De ahhoz, hogy egyáltalán megértsék egymást, hogy a problémát közösen legyenek képesek megoldani, az kell, hogy a társadalomtudós értsen valamit a matematikához. Másrészt a matematikusnak is bizonyos érzéke kell, hogy legyen a kérdésfeltevő tudományához. Enélkül semmire sem jutnak. Pedig egymás megértése nagyon fontos, különösen a kapott eredmények értelmezésénél. Sem a matematikától, sem a számítógéptől nem várhatjuk, hogy a felvetődött társadalmi problémát megválaszolja. A kérdést a kutató teszi fel, s ugyanúgy ő adja a

választ is, ha tudja. A számítás eredményét, a gép outputját a kutató kell, hogy értelmezze. Felfogásom szerint matematika alkalmazásával a társadalomról megfogalmazott állítás se igazabb, se egzaktabb nem lesz. Bizonyításként is csak akkor fogadhatjuk el, ha a kinduló adatokat elég megbízhatónak és a módszert jogosultnak találjuk. Ezek nélkül matematikai módszerek használata csak bűvészkedés a számokkal. Hozzátehetjük, nagyon sok esetben – jobb híján – olyan eljárást kényszerülünk alkalmazni, amelynek még a matematikai egzaktsága sem bizonyított.

Hogyan lehet elképzelni a történész és a matematikus együttműködését? Hát ez elég problematikus. Úgy gondolom, hogy a matematikus rendszerint egyszeri megbízással, pénzért végzi. Bizonyos idő után kialakul a piacon egy olyan matematikus-programozó társaság, amelyről tudni lehet, hogy ért ezekhez a dolgokhoz.

Granasztói: A számítástudományi szakember azonos egy programozóval?

Göndör: Nem feltétlenül. A programozó középszintű végzettséggel rendelkező szakember, aki algoritmus alapján megír egy gépi programot, futtatja, intézi a vele kapcsolatos ügyeket. A probléma matematikai modelljének megtalálása, és ennek alapján egy programozható algoritmus kidolgozása viszont a matematikus feladata. Nem túl bonyolult problémák esetén nem baj, ha a két, egyébként világosan különböző funkciót egy ember tölti be. Tévedés lenne tehát azt képzelni, hogy a számítógép alkalmazása kizárólag a programírából áll. Egyáltalán nem hiábavaló, ha nagy vonalakban végig gondoljuk a szükséges munkafolyamatokat. A munka az adatelőkészítéssel kezdődik. A probléma matematikai modelljével összhangban ki kell jelölni a rendelkezésre álló forrásadatokat az azt a körét, amelyeket a kutatásba be kívánunk vonni. A következő fázis az adatok gépi adathordozóra való felvitele. Itt meg kell tervezni a formát, amely ugyancsak függ attól, milyen módon kívánjuk majd a továbbiakban feldolgozni. Az adatrögzítést az ellenőrzés és javítás addig ismételt ciklusai követik, amíg az anyag a kívánt hibátlanúsági fokot el nem éri. (Szociológiai kérdőíves felvételek feldolgozásánál ez a fázis sokszor kimarad. Néhány százaléknyi hiba a legtöbb statisztikai módszert lényegében nem zavarja. De egy régészeti adatnyilvántartásban az elvárások szigorúbbak.)

Granasztói: Hogyan kezeli az adatokat a SZOCPROG-rendszer?

Göndör: A SZOCPROG már a következő munkafázissal foglalkozik, vagyis a már előkészített adatok statisztikai feldolgozásával. Az adatok feldolgozása – ha a probléma jellege ilyen – történhet standard matematikai-statisztikai eljárások valamelyikének alkalmazásával. Erre a SZOCPROG bizonyára nagyon jó.

Granasztói: Itt döntő különbség van – azt hiszem – a tudományágak, például a szociológia és a történettudomány természete között. Ez mutatkozik meg most.

Göndör: Elképzelhető, hogy a történész is tud használni standard matematikai-statisztikai módszert. Vannak történeti-statisztikai adatok: népszámlálások, különféle összeírások. Ezeknél lehet értelme statisztikai módszereknek, de lehet ezer más eset, ahol nincs értelme, mert az adatok nem is kvantitatívak, hanem pl. szövegesek, amelyek kvantifikálása nem is értelmes, mert esetleg maga a szöveg az érdekes, a feladat ezeknek a szövegeknek a pusztá nyilvántartása és rugalmas több szempontú visszakeresése, folyamatos javítása, kiegészítése.

De térjünk vissza ahhoz a kérdéshez, hogy tervezheti meg az ember a kutatását. A szociológus helyzete ebből a szempontból, bizonyos mértékig, könnyebb. Van egy társadalmi problémája, ehhez adatokat tud gyűjteni, méghozzá olyanokat, amelyeket a rendelkezésre álló adatfeldolgozási technikával kezelni is tud. A történésznek nincs ilyen lehetősége. A forrása adott és ezt vagy tudja a meglévő technikájával kezelni, vagy nem.

A szociológiában elterjedt a véletlenszerű mintán alapuló kérdőíves interjútechnika. Ilyenkor az egyes interjúk egy-egy adategységnek felelnek meg – mondjuk, egy rekordnak – és éppen a minta véletlenszerűsége miatt, az adategységek között nem vizsgálják a kapcsolatot. Ez az oka annak, hogy a szociológus a lehető legegyszerűbb adatszerzési módszerrel is boldogul. Egy történésznek ilyen szerencséje ritkán adódik. Ha például háztartásokat vizsgál, nyilván akarja tartani a rokoni viszonyokat, szomszédi viszonyokat stb. A mintája sokszor szórványos, amit viszont teljeskörűen akar elemezni. Tehát a véletlen minta nyújtotta előnyöket nem tudja kiaknázni, így magára a struktúrára koncentrálni. Bonyolult adatstruktúrával dolgozik, amelynek a gépi megvalósítása is általában bonyolult. Nagyobb és korszerűbb gépeken ezért ilyen célt szolgáló programrendszerek már általában készen léteznek.

Még egy problémára szeretném felhívni a figyelmet. Arra, hogy a társadalomtudományi kutatók munkahelyein egyrészt a gazdasági szabályozók, másrészt a határidők rendszere miatt csak a legritkább esetben kerülhet sor arra, hogy egy gépes feldolgozásra szánt munka matematikai modelljét valóban matematikus dolgozza ki. Legtöbbször a kutatás végén kapcsolódik be egy programozó, aki így kész tények elé kerül, ráadásul őt az elkészített táblák száma szerint fizetik. A munka rendszerint sürgős, a programozó elemi érdeke tehát nem az, hogy új módszereket keressen, hanem, hogy egy-két meglévő standard módszert alkalmazzon.