

AZ ÖSSZEÍRÁS DIÓHÉJBAN



2. rész: Egy kis adatgyűjtés-módszertan



2014. január

Tartalomjegyzék

| | |
|--|----------|
| Tartalomjegyzék..... | 2 |
| Előszó..... | 3 |
| V. Adatgyűjtési módszerek egy- és több csatornán | 4 |
| V.1. Az összeíró által végzett személyes kérdezés: papíron és számítógépen | 4 |
| V.2. A telefonos lekérdezés | 5 |
| V.3. Válaszadás önállóan, összeíró nélkül: papíron és interneten..... | 6 |
| V.4. Többcsatornás adatgyűjtések | 7 |
| VI. Teljes körű és mintavételes adatgyűjtések | 8 |
| VI.1. Teljes körű adatgyűjtések | 8 |
| VI.2. Mintavételes adatgyűjtések..... | 9 |
| Mintavételi eljárások | 10 |

Előszó

A Központi Statisztikai Hivatal adatfelvételeiben közreműködő összeírók felelőségteljes feladatot látnak el, és egyfajta „híd” szerepet töltenek be az adatszolgáltatók és a hivatal között. Felkészültségük, szaktudásuk, a válaszadókkal való együttműködésük minősége jelentős mértékben meghatározza az adatgyűjtések sikerét, eredményességét.

Az összeíróknak nem csak a **kérdőívek kitöltésében** kell jártasnak lenniük. Amikor válaszadásra kérik a kijelölt címeiken élő személyeket, vagy egyéb módon (pl. bolti árak megfigyelésével) adatot gyűjtenek a statisztikai felmérésekben, **meg kell győzniük az adatszolgáltatókat** arról, hogy fontos a részvételük, szakszerűen tájékoztatniuk kell őket az adatgyűjtés céljáról, módszeréről. Emellett az adatszolgáltatók is számtalan kérdést intézhetnek feljük, többek közt az adatgyűjtés módszerére, technikájára vonatkozóan. Az összeírók felkészülésében tehát szerepet kell kapjanak olyan általánosabb, a statisztikai adatfelvételekkel kapcsolatos tudnivalók is, amelyek munkájukat érinthetik.

Mostani kiadványunk célja, hogy rövid áttekintést nyújtson az összeírók **alapozó felkészüléséhez** a főbb adatgyűjtési eljárásokról, a személyes kikérdezéstől, a telefonos interjún át a válaszadók által önállóan történő internetes (online) kérdőív-kitöltésig, valamint a papír alapú és elektronikus adatgyűjtő eszközökről. A leendő és a kezdő összeírók bepillantást nyerhetnek a **mintavétel** rejtelseibe is.

V. Adatgyűjtési módszerek egy- és több csatornán

Ha valamilyen döntés vagy fejlesztés megalapozásához információra van szükség (pl. egy internethálózat bővítéséhez), akkor ezeket valamilyen módon össze kell gyűjteni. Ezt az adatgyűjtést többféle formában végezhetjük el. Kielégíthetjük az információigényt úgy, hogy már rendelkezésre álló **dokumentumokból, nyilván-tartásokból** válogatjuk ki a szükséges adatokat (pl. lexikonok, lakcímnnyilvántartás, cégjegyzékek, adminisztratív nyilvántartások vagy az internet-előfizetők jegyzékéből). Más esetben pl. **megfigyelhetjük** a vizsgálandó jelenséget (pl. a vásárlási szokásokat), tárgyakat, embercsoportokat, szervezeteket. Bizonyos kérdésköröknél azonban leghitelesebb az adatforrás, ha **közvetlenül személyesen** keressük meg és kérdezzük ki az érintetteket. Ezt az adatgyűjtést egyfelől végezhetik **kérdezők, összeírók** – papír alapú vagy számítógépre telepített kérdőíven –, másfelől a **cél-személyek** (vagy intézmények) **maguk is megválaszolhatják** a kérdéseket (postán eljuttatott papír kérdőíven vagy elektronikus úton, pl. az interneten keresztül).

Az **adatgyűjtési mód** megválasztásánál azt mérlegelik a statisztikusok, hogyan lehetne leghatékonyabban, leggyorsabban és legmegbízhatóbban megszerezni az információkat az érintettektől. Mi a célravezetőbb? Egy közvetlen megkeresés, személyes beszélgetés (interjú) a válaszadóval? Vagy jobb egy személytelenebb, közvetett kapcsolat telefonon keresztül? Vagy inkább az a célszerű, ha a kérdezett nem is kerül kapcsolatba az adatgyűjtővel, és saját maga tölti ki a kérdőívet? A különböző adatgyűjtési eljárásoknak eltérőek az előnyei, és bizonyos hátrányaikkal is számolni kell, amikor valamelyik mellett döntenek. A következőkben a leggyakrabban alkalmazott kérdőíves adatgyűjtési módszerek jellemzőit vesszük sorra.

V.1. Az összeíró által végzett személyes kérdezés: papíron és számítógépen

Az **összeírók, kérdezők közreműködésével zajló személyes lekérdezés (face-to-face interjú)** a kérdőíves felmérések leggyakoribb módszere mindmáig, mivel más eljárásokhoz képest lényegesen alacsonyabb válaszmegtagadást eredményez. Ennek oka, hogy egy összeíró által elhangzó udvarias, **személyes kérdésnek** nehéz ellenállni. Másfelől a válaszadók jobban szeretik, ha valaki **segíti** a kérdőív kitöltését, az útmutatások betartását, a kérdések és válaszok értelmezését, és nem maguknak kell mindezzel bajlódniuk. Ugyanakkor a kérdező pozitív szerepét beárnyalja az, hogy **jelenléte befolyással** lehet a válaszokra: pl. érzékenyebb, illetve véleményre irányuló kérdések esetén neki esetleg kevésbé adnak őszinte válaszokat. Ilyenkor az is előfordulhat, hogy valamilyen elvárásnak próbálnak megfelelni a megkérdezettek.

Mindent egybevetve, a személyes kikérdezés eredményességében **rendkívül nagy szerepe van a kérdező személyének**, külső megjelenésének, szavakkal, magatartásával, testbeszédével kommunikált üzeneteinek, amelyeket a válaszadó felé közvetít. Ha mindez kedvező benyomást gyakorol a megkeresett személyre, akkor nagy valószínűséggel hajlandó lesz az együttműködésre.



Napjainkban a **papír kérdőívek** egyre inkább háttérbe szorulnak, és helyette számítógép segíti az összeírók munkáját. A **számítógéppel támogatott személyes interjúban** egy mobil elektronikus eszközbe, pl. **laptopba** rögzíti az összeíró a válaszokat. Ez a megoldás jelentősen csökkenti a költségeket, mivel itt a **kérdezés és**

A kérdezővel segítségével zajló adatgyűjtés biztosítja a legnagyobb válaszadási késziséget.

az adatrögzítés folyamata egybeesik. Ez a megoldás a papír kérdőívvel szemben jelentősen megkönnyíti az összeíró munkáját is, az alábbiak miatt:

- A sok papír kérdőív **cipelése** helyett csak egy hordozható számítógépet kell az összeírónak magával vinnie.
- A számítógép **automatikusan** jeleníti meg a soron következő megválaszolandó kérdéseket, a válaszadó személyére szabottan, ezért az összeírónak nem kell keresgélnie, hogy melyik kérdéssel folytassa a beszélgetést, vagyis nem kell figyelnie az ún. „**ugratásokat**”, bizonyos kérdések logikai szempontból való kihagyását. (Például ha a válaszadó nőtlen/hajadon, akkor nem teszünk fel neki kérdéseket a házastársáról; vagy ha gyermektelen, akkor ki kell hagyni a gyermekekre vonatkozó kérdéseket.)
- **Gyorsabb** a válaszok, válaszlehetőségek rögzítése, mert a kérdező egy listából (pl. országnevek listája ABC-sorrendben) jelölheti ki a lehetséges válaszokat, így nincs szükség kézírásra, jegyzetelésre.
- A számítógép figyelmezteti az összeírót, ha **tévesnek** tűnő adatot próbál bevinni (pl. jelzi, ha a válaszadó beírt születési dátuma túl magas életkort ad ki)
- **Logikai összefüggéseket** is vizsgál és figyelmezteti az összeírót (pl.: ha 20 éves a válaszadó és az összeíró azt rögzíti, hogy van unokája, akkor a program visszakerdez, hogy biztos-e a válasz).
- A „**súgó**” **segítő funkció** révén könnyebb az útmutatások kezelése, szabályok betartása, mint egy papírra nyomtatott többoldalas útmutatóval.
- Ahol **többször kell visszamenni** egy adatszolgáltatóhoz (pl. a munkaerő-felmérésben), ott a korábbi adatgyűjtésből származó információkat **nem kell újra megkérdezni**, mert ezek egy része automatikusan megjelenhet a kérdőívben.
- A válaszadók úgy érezhetik, hogy az adataik nagyobb **biztonságban** vannak a számítógépen, mint papíron.

A számítógéppel támogatott lekérdezés tehát egyfelől könnyebbséget jelenthet, ugyanakkor az **eszközhasználat, a technikai problémák** olykor nehézséget okozhatnak, ezért nagy gyakorlatot kell szerezni ahhoz, hogy a kérdezés zökkenőmentesen történjen, és az összeíró esetleges gyakorlatlansága ne növelje a válaszadó türelmetlenségét. Ezért a kérdezőknek a felkészülés során ún. **"gyakorló üzemmódban"** lehetőségük van és fontos is megismerni az elektronikus kérdőív kitöltésének technikáját.

V.2. A telefonos lekérdezés

A telefonos lekérdezés előnye a nagyobb személytelenség

A telefonos kikérdezés előnye, hogy itt is érvényesül az **összeíró személyes meggyőző ereje, és felkészültségével** segíti a kérdések megválaszolását. Ugyanakkor a telefonos interjú biztosította **személytelenség** könnyebben elérhetővé és őszintébbé teheti a „rejtőzködő” adatszolgáltatókat, ily módon csökkentheti a válaszok torzítását és a válaszmegtagadás esélyét. Ez a személytelen, közvetett kapcsolat az adatszolgáltató számára kényelmesebb (akár pizsamában is válaszolhat), és kevésbé feszélyező, mint a személyes találkozás. Ezért kényesebb, bizalmasabb témák esetén is jól alkalmazható. Amennyiben a válaszok lejegyzése **számítógépbe** történik, akkor természetesen érvényesülnek az elektronikus kérdezés előnyei is.

Ugyanakkor a telefonos adatgyűjtés általában na-



nagyobb arányú **válaszmegtagadással** jár, mint a személyes (face-to-face) beszélgetés, ugyanis telefonon keresztül sokkal könnyebben utasítanak vissza, mint szemtől szemben állva. A telefonos adatgyűjtéseknél jóval kevesebb – lehetőség szerint maximum 10 perces – lehet csak az adatfelvétel **időtartama**. Ennek következtében a kérdésekre adható **válaszlehetőségek** száma is csak korlátozott lehet.

Az összeírónak másfajta készségekre van szüksége telefonos adatgyűjtés során, mint egy személyes interjúban, ahol leginkább a jó megjelenés és kedves, határozott fellépés a siker titka. A telefonos kérdezés esetében inkább a hangszín, a hangsúlyozás, a „**mosolygós hang**” a benyomás alapja, ezért nem könnyű elérni, hogy telefonon keresztül is meggyőzőek tudjunk lenni.

A KSH néhány olyan adatfelvételében, ahol **több alkalommal** is vissza kell menni az adatszolgáltatóhoz (pl. munkaerő-felmérés), az összeíró a **második felkeresést már telefonon** végezheti. Vannak olyan korszerű megoldások is, ahol a kérdezők egy központban, ún. **call centerben** ülve kérdezik le a válaszadókat, általában automatikus tárcsázás lehetőségével. Ekkor a számítógép maga hívja a berögzített telefonszámokat, úgy, hogy az operátorhoz csak azok a hívások kapcsolnak be, amelyeket fel is vettek. Ebben a rendszerben az is beállítható, hogy egy számot egy adott időpontban hívjanak fel, ami olyankor hasznos, ha a válaszadó nem ér rá, de szívesen válaszolna egy másik időpontban, vagy ha a lekérdezést valamiért félbe kellett szakítani.

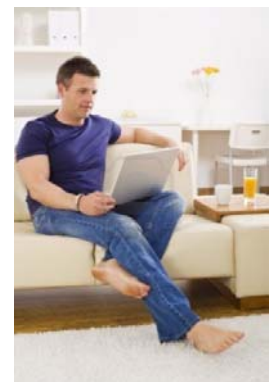
V.3. Válaszadás önállóan, összeíró nélkül: papíron és interneten

Az önkitöltős kérdőív a legkényelmesebb a válaszadók számára.

Ma már egyre jobban igénylik a válaszadók, hogy **saját maguk**, a nekik tetsző időpontban és helyszínen, akár többszöri megszakítással is, függetlenül a kérdező jelenlététől, és akár „pizsamában” is megválaszolhassák a kérdéseket. Míg korábban ezt **postai** úton kijuttatott vagy a kérdező által átadott **papír** kérdőíveken tehették meg, addig ma ugyanezt még egyszerűbben, számítógépen, az **interneten**, weben keresztül is.

Az **önállóan megválaszolt ún. „önkitöltős” papír kérdőíveknek** (amit pl. a népszámlálásnál is alkalmaztak) sok hátrányát ismerjük. Nehézséget okoz a **válaszadók írásának**, jelöléseinek olykor nehéz olvashatósága, rögzíthetősége, a válaszadóra nem vonatkozó, fölöslegesen kitöltött kérdések, vagy éppen fordítva: a hiányos kérdőívek, továbbá az útmutatások figyelmen kívül hagyása miatt a félreértelmezett kérdésekből adódó használhatatlan válaszok. A papír kérdőívek igen megnehezítik a **feldolgozást** is, mert sok utómunkálatra (pl. a nyitott válaszok leködölésére van szükség).

Az **elektronikus, illetve internetes (webes, online) önkitöltős kérdőívek** azonban többnyire kiküszöbölik ezeket a problémákat, mert – hasonlóan az összeíró által használt elektronikus kérdőívhez – automatikusan **beépített útmutatók** figyelik, megfelelő kérdéseket tölt-e ki a válaszadó, a beírt válaszok értelmezhetőek-e, valószínűsíthetőek-e, stb. A válaszadó nyomon követheti hol tart a kitöltésben, visszalapozhat, könnyen javíthatja a válaszait.



Az internetes kérdőív olcsó, és segíti a kitöltést – de kevesebben válaszolnak rá.

Az **internetes** adatgyűjtés nagy előnye, hogy – amint a számítógéppel támogatott kérdezői interjúnál is – itt is egy időben történik a **válaszadás és az adatrögzítés**, ami lehetővé teszi az adatok gyors feldolgozását, részeredmények közlését. Az internetes adatgyűjtés lényesen

olcsóbb, mint a hagyományos papírkérdőíves lekérdezés, mivel nincs szükség utazásra, kérdőívek nyomtatására, telefonra, laptopokra, stb.

Néhány **hátránya** azonban az internetes önkitöltésnek is adódik, ami elsősorban magából az adatszolgáltatók önálló, **kérdező segítsége nélkül** történő válaszadásából, a kérdések, válaszlehetőségek nehezebb vagy hibás értelmezéséből fakad. Hasonlóan a postán kiküldött papír kérdőívhez, itt is nagyon nehéz elérni, hogy a válaszadók **határidőre** vagy egyáltalán kitöltsék a kérdőívet, mert **nem motiválja** őket az összeíró személyes meggyőző ereje.

Mindezért igen alacsony **válaszadási aránnyal** kell számolni, vagy nagy energiát kell fordítani a beszüretésre, meggyőzésre. Emellett még ma sem rendelkezik minden háztartás internettel, és nem minden adatszolgáltató tudja/akarja kezelni a számítógépet. Ezért általában ennek a módszernek az egyedüli alkalmazása nem elegendő, hanem **más válaszadási lehetőségeket is** fel kell kínálni az érintetteknek.

V.4. Többcsatornás adatgyűjtések

A többcsatornás adatgyűjtés a válaszadó igényeit szolgálja.

Az egyes adatgyűjtési módszereket célszerű **kombinálni** annak érdekében, hogy mindegyiknek az előnyét érvényesíthessük, és csökkentjük a hátrányuk okozta problémákat, továbbá hogy az adatszolgáltató egyéni igényeihez igazodó megoldást kínálhassunk. Mikor egy felmérésen belül többféle válaszadási módot (pl. internetes önkitöltést, telefonos megkeresést vagy összeíró segítségével történő személyes válaszadást) is alkalmazunk, akkor **többcsatornás, vegyes módszerű vagy ún. „hibrid” adatgyűjtésről** beszélünk.

A többcsatornás adatgyűjtés növeli a válaszadói hajlandóságot.

A leggyakoribb megoldás, amikor a **megkérdezettekre bízzák a válaszadás módját** azzal a céllal, hogy a számára legkényelmesebb megoldással növeljék a válaszadói hajlandóságot, motiváltságot, és ezáltal az adatok minőségét is. Feltételezhető ugyanis, hogy bizonyos csoportokat könnyebb az interneten elérni, mint személyesen (pl. a fiatalok vagy a dolgozók napközben kevéssé elérhetők, míg a nyugdíjasok könnyebben). A 2011-es Népszámlálásnál a háztartások eldönthették, hogy miként válaszolnak. Végül majdnem minden ötödik címen az interneten keresztül online kérdőívet töltöttek ki, de a legtöbben (a háztartások közel kétharmada) a számlalóbiztosok segítségét vette igénybe. Ennél sokkal kevesebben saját maguk álltak neki a papír kérdőívek kitöltésének.



Egy másik megoldás, hogy **ugyanannak a kérdőívnek** egy nehezebben megválaszolható részét **összeíró** közreműködésével kérdezik le, míg az egyszerűbb vagy éppen kényesebb kérdéseket **maguk a válaszadók** tölthetik ki.

A többcsatornás adatgyűjtéseknek további előnye a **költséghatékonyság**. Ha a személyes lekérdezés mellett a kérdezettek egy része pl. telefonon vagy interneten válaszol, akkor utazási költségeket spórolhatunk meg, és telefonos lekérdezés esetén a távolságok sem jelentenek problémát, az összeíróknak nem kell a címek felkeresésével időt töltenie, és az ország bármelyik részén élő lakossal készíthet interjút.

A többcsatornás adatgyűjtés egyik hátránya a „módhatás”.

A többcsatornás adatgyűjtésnek számos előnye mellett **hátrányaival** is számolnunk kell, többek közt azzal, hogy az adatgyűjtés **megszervezését** nagymértékben bonyolítja, hiszen a különböző módon beérkező, más-más formájú (papír, laptop,

internet) kérdőívekből egy közös csatornába kell terelni az adatokat. Más oldalról felvetődik az a kérdés is, „*vajon ugyanaz a személy ugyanazokat a válaszokat adná-e, ha egy másik módon kérdeznék le*”. Ugyanis nem hagyható figyelmen kívül, hogy a több csatornán beérkező válaszok különbségeinek egy része abból adódik, hogy az egyes megkérdezettek **más-más szituációban válaszoltak** ugyanazokra a kérdésekre (pl. az egyikük egyedül otthon, a másik az összeírónak személyesen), ami befolyásolhatja a hozzáállásukat, végül válaszaikat is. Ez az ún. „**módhatás**”, ami leginkább a nem tényszerű információkat, az érzékeny témákat vagy a véleménykérdéseket érintheti (pl. az alkoholfogyasztásról őszintebben válaszolnak a megkérdezettek, ha maguk töltik ki a kérdőívet, mint amikor egy összeíró faggatja őket).

VI. Teljes körű és mintavételes adatgyűjtések

A válaszadók gyakran kérdezik: „*Hogyan választottak ki éppen engem?*”

A válaszadók egyik leggyakoribb kérdése az összeírók felé, hogy „*Miért éppen nekem kell válaszolni? Hogyan választottak ki? Miért nem inkább a szomszédomat kérdezi le?*” (erről röviden szóltunk az összeíró tájékoztató feladatát bemutató II.6. fejezetben). Ha az összeíró eredményesen szeretné végezni a munkáját, kielégítő választ kell tudnia adni az ilyen és hasonló kérdésekre. Ehhez pedig némi ismeret-re van szüksége az adatfelvételek mintaválasztási eljárásáról.

Vannak olyan adatgyűjtések, amelyek a vizsgálni kívánt jelenség vagy célcsoport teljes körére irányulnak, és valamennyi egyedet számba veszik. Más esetben azonban nem tudjuk, vagy nem is akarjuk megkeresni a célcsoport egészét, ezért csak egy kisebb részét, mintáját választjuk ki, ügyelve arra, hogy a kiválasztott egyedek megfelelően képviseljék (reprezentálják) a teljes sokaságot. Az összeírók – eltekintve a népszámlálástól – többnyire ilyen mintavételes adatgyűjtésekben vesznek részt.

VI.1. Teljes körű adatgyűjtések

Teljeskörű adatgyűjtésre van szükség alapstatisztikák létrehozásához.

Bizonyos jelenségek feltárása, értelmezése csak úgy lehetséges, ha annak teljességéről van információnk. Az, hogy egy adott évben hányan és mikor, hol **születnek**, valamint hányan és milyen betegségeken **halnak** meg, vagy hány **házasságot** kötnek, és hányan **válnak** el, csak **valamennyi idevágó eset számbavételével** deríthető fel. Ezekről a jelenségekről teljes körű adatgyűjtéseket kell végrehajtani. Szintén valamennyi intézmény bevonására van szükség például az **iskolák** adott évben látogatott szakjairól, hallgatói **létszámáról**, vagy az ország különböző településein működő kórházak **betegellátásáról** szóló információk megszerzéséhez. Nehéz lenne ezeket az adatokat **becslések** útján előállítani, mivel ezek **alapstatisztikák**.



Vannak olyan adatgyűjtések, amelyeknek az a célja, hogy **pontos alapadatokat szolgáltatassanak** döntésekhez, vagy további adatok, elemzések kiindulópontjaként szolgáljanak. Ilyen megfontolásból is tartanak rendszeres időközönként a világ szinte valamennyi országában nép- és lakásszámlálásokat vagy pl. mezőgazdasági cenzusokat. Az ezek során teljeskörűen (minden egyes településen mindenkiről)

gyűjtött adatok megbízható alapul szolgálnak a további vizsgálatokhoz, kutatásokhoz, és ezek jelentik a **mintavételes adatgyűjtések kiinduló alapstatisztikáit** is.

VI.2. Mintavételes adatgyűjtések

A teljes körű megfigyelések mellett a statisztikai adatgyűjtések másik nagy csoportját a mintavételen alapuló eljárások alkotják. Ezekre általában a "**reprezentatív megfigyelések**" kifejezéssel hivatkozunk. Napjainkban a médiában igen gyakran találkozunk a közvélemény-kutató intézetek beszámolóival, amelyek ezerkétezer személy megkérdezése alapján meglepően jó képet adnak a társadalmat foglalkoztató kérdések fogadtatásáról. Ezek úgynevezett reprezentatív megfigyelések, amelyek esetében egy általában kis részből, **a mintából következtetünk az egészre**. Amikor valaki otthon pl. levest főz, akkor elég egy kanalat kellő mélyen a lábosba meríteni ahhoz, hogy meg tudja állapítani az ízet vagy a zöldségek puhaságát.

Mintából következtetünk az egészre.

A statisztikai adatgyűjtések esetében az „egészet” valamilyen adott terület (pl.: ország, régió) lakóinak, háztartásainak vagy éppen lakásainak az összessége jelenti. Ezekből veszünk megfelelően kiválasztott, az **egészet jól képviselő (reprezentatív) mintát** ahhoz, hogy a teljes sokaságra vonatkozóan minél pontosabb becsléseket tudjunk megfogalmazni. Ezeknek az úgynevezett reprezentatív felméréseknek az a nagy előnye, hogy sokkal kevesebb költséggel és szervezéssel járnak, mint a teljes körű adatfelvételek.

Az adatgyűjtéseknek két fő **célcsoportja** van:

- A **személyes lekérdezést, ún. interjúk módszert alkalmazó kérdőíves** adatgyűjtések az esetek döntő részében a **lakosság**, illetve a lakosság egy meghatározott rétege, csoportja körében folynak, vagyis az összeírók többnyire a lakossággal (ezen belül **egyénekkel vagy családokkal**) kerülnek kapcsolatba.
- Vannak azonban olyan feladatok, amikor a kérdezők, összeírók nem magánszemélyek, családok körében végzik az adatgyűjtést, hanem **intézményekben, gazdasági szervezetekben**, boltokban, elárúsító helyeken.



A minta segítségével becsüljük a sokaság jellemzőit.

A **mintavételes adatgyűjtéseket**, összeírásokat azért részesítik előnyben, mert a vizsgálni kívánt célcsoport egészét, teljes körét nem lehet terhelni gyakori felkereséssel, kérdezéssel, és ez anyagilag is rendkívül költséges lenne. De nincs is szükség erre, hiszen vannak olyan **matematikai-statisztikai módszerek**, amelyekkel egy országos címjegyzékből (vagy pl. boltok összeírása esetén különböző szakmai jegyzékekből) **véletlenszerűen ki lehet választani** a vizsgálni kívánt csoportot jól képviselő arányos részt, egy úgynevezett reprezentatív mintát. A tőlük összegyűjtött adatok alapján pedig **becslések** útján következtetni lehet a teljes sokaságra.

A minták segítségével a **sokaság különféle jellemzőit** lehet becsülni. A becsült adatok leggyakrabban a következők lehetnek:

- **létszámok** vagy értékösszegek (pl. foglalkoztatók létszáma, munkanélküliek létszáma),



- **átlagok** (pl. adott korcsoporthoz tartozó személyek átlagos testsúlya vagy magassága egy egészségügyi felvételen),
- **arányok** (pl. a munkanélküliek aránya a gazdaságilag aktív népességben).

A minta csak egy része a sokaságnak.

Mivel a **minta csak egy része a sokaságnak**, ezért a mintából származó becslések általában nem egyeznek meg azzal az értékkel, amit akkor kapnánk, ha mindenkit megkérdeznénk. Ezt az eltérést nevezzük **mintavételi hibának**, ami a mintavételi eljárás bizonyos mértékű pontatlanságának következménye lehet. Például olyanok is bekerültek a mintába, akik nem tartoznak az érintett célcsoportba (pl. egy fiatalokat vizsgáló felmérés esetén egy 60 éves válaszadó kerül be a mintába), vagy éppen hiányzik egy bizonyos oda tartozó csoport (pl. a 20 évesek). Az is mintavételi hibának számít, ha pl. a valóságosnál nagyobb arányban képviselnek egy bizonyos csoportot (pl. ha túl sok nyugdíjas válaszol a kérdésekre, akkor az ő állás-pontjuk túlsúlyba fog kerülni).

Mintavételi eljárások

A véletlen befolyásolja, hogy melyik cím kerül be a mintába.

A KSH-ban csak úgynevezett **valószínűségi mintákat** alkalmaznak, amelyek lényege, hogy az alapsokaság minden elemének (pl.: a lakosság minden személyének vagy háztartásának) van egy meghatározott esélye arra, hogy bekerül egy adatgyűjtésbe, felmérésbe. Ez azt jelenti, hogy a véletlen befolyásolja, hogy egy háztartás vagy egy személy kiválasztásra kerül-e, vagy sem; az emberi tényezőnek – a matematikusnak, az adatgyűjtés-szervezőnek vagy az összeíróknak – azonban nincs szerepe a válaszadók kiválasztása során.

A legegyszerűbb és egyben elméleti szempontból a legjobb tulajdonságú az ún. **egyszerű véletlen mintavétel**. Szemléletesen a következőt jelenti: a jelenleg kb. 10 milliós magyar lakónépességből 1000 főt úgy választunk ki **egyszerű véletlen mintavétel** esetén, hogy a 10 millió személy nevét felírjuk egy cédulára, bepakoljuk egy kalapba, és kihúzzunk belőle 1000 darabot. Ezt **egylépcsős mintavételnek** nevezzük.



A rétegzett mintavételben egységes jellemzők szerinti csoportokból választunk adatszolgáltatókat.

Ez a módszer azonban sokszor vagy nem alkalmazható, vagy az adatgyűjtés végrehajtásának gyakorlati szempontjaiból nem kívánatos. Például összeíró közreműködésével végrehajtott adatgyűjtés esetén előfordulhatna, hogy a kiválasztott 1000 cím nagyon sok településen szóródna szét, és ezek felkeresése rendkívül sok időt és költséget jelentene. Egyszerű véletlen minták helyett ezért valamilyen szempontból csoportosított mintákat alkalmaznak. Ilyen például a **rétegzett mintavétel**, amelyben a vizsgált alapsokaságot bizonyos szempontok szerint egységes jellemzőkkel rendelkező csoportokra bontják (pl. a települések lélekszáma alapján). Ezt követően ezekből az egységes csoportokból (pl. az 5-10.000 fős településekből) választanak ki véletlenszerűen néhányat (néhány ekkora települést), és utána már csak ezeken a településeken vesznek mintát a háztartásokból. Az adatfelvételek egy részében a kiválasztott háztartásokból csak egyetlen, valamilyen módszerrel kiválasztott személyt kérdeznek le, másik részükben pedig a háztartás valamennyi tagját.

