

# A mintavételi tervek készítésének néhány gyakorlati megfontolása

*A gyakorlatban a közvéleménykutató cégek, a különböző elégedettségvizsgálatok, pártpreferencia-felmérések és még számos – társadalomtudományi indíttatású – terület szinte „naponta” kényszerülnek mintavételre, reprezentatív minta alkalmazására. Az elméleti szakemberek számára felettébb bosszantó, hogy ezek a – sokszor nagy fontosságú – kérdések módszertani szempontból mennyire nem átgondoltan, sokszor „pongyolán” intéződnek; milyen gyakori, hogy szakmailag részben vagy teljesen kifogásolható eredményeket közölnek.*

*A gyakorlati mintavétel, illetve mintavételi tervekészítés legtöbbször három területen ütközik az elmélettel, illetve három jelentős kérdés merül fel:*

- Milyen is a reprezentatív minta?
- Mekkora legyen az alkalmazott minta?
- Milyen mintavételi módokat alkalmazzon a felhasználó?

*A cikk e három kérdést érinti vázlatosan, főképpen a gyakorlatban sokszor elkövetett hibák, problémák oldaláról szemlélve azokat.*

## Milyen a reprezentatív minta?

A magyar statisztikai szakirodalomban mindmáig alapműnek számító „Köves – Párniczky” szerint<sup>1</sup> a reprezentatív mintavételnek az alábbi sajátosságai vannak:

- A mintavétel véges sokaságból történik, az alapsokaság tömeges (legalább ezres nagyságrendű);

1 Köves Pál – Párniczky Gábor: Általános statisztika I-II. KJK, Budapest 1981., 103–104. old.

2 A magyar statisztikai szóhasználatban a reprezentatív megfigyelés, mint a mintavétel gyakori módszere Párniczky Gábor – Csepinszky Andor: *Reprezentatív megfigyelés a gazdaságstatisztikában* című könyvének megjelenése (KJK 1956) óta terjedt el.

- A tömeges sokaságból általában nagy mintát veszünk (a minta-elemszám legalább 100);
- A mintavétel általában visszatevés nélkül történik, így a mintánk nem tekinthető független azonos eloszlású (FAE) mintának;
- Az így létrejött mintából származó információt általában becslés céljára használjuk (hipotézisvizsgálatot csak ritkán alkalmazunk, hiszen a tesztek általában FAE mintát igényelnének);
- A következtetési statisztika torzítása nem csak a becslőfüggvény eloszlásából ered, hanem nagymértékben származik a mintavétel módjából;
- A minta kialakításában fontos szerepet játszanak gyakorlati szempontok, pl. a sokaság térbeli elhelyezkedése, a mintavétel költségei, a címlisták megléte stb.

A fenti módon körülírt, véletlen eljárással nyert minta akkor „reprezentatív”, ha képviseli a sokaságot. A mondat komoly veszélyeket rejt magában, ha a felhasználók nem kellően gondolják át a tartalmát. A mintába kerülő egységeket valóban elfogulatlanul, az önkényesség kizárásával kell kiválasztani, amelyek így reprezentálják, képviselik az alapsokaságot. Maga a kiválasztás, a megfigyelés ezt a reprezentativitást igyekszik biztosítani, elterjedt szóhasználatlál élve reprezentatív megfigyelést<sup>2</sup>, kiválasztást végzünk. A végeredmény, a minta azonban gyakorta nem reprezentatív.

Reprezentatívnek tekinthetjük azt a mintát, melynek legfontosabb jellemzői nem térnek el lényegesen a sokaság megfelelő jellemzőitől. Mindez azt is jelenti, hogy a reprezentativitás csupán viszonylagos lehet, függ a fontosnak tartott sokasági jellemzőktől. Teljesen reprezentatívnek akkor mondhatnánk a mintát, ha az valamennyi elemében, összetételében és fő jellemzőjében megfelelne az alapsokaságnak. Ter-

”

A véletlen eljárással nyert minta akkor „reprezentatív”, ha képviseli a sokaságot. A mondat komoly veszélyeket rejt magában, ha a felhasználók nem kellően gondolják át a tartalmát.

”

mészletesen ez utóbbi esetben már nem mintáról, hanem magáról az alapsokaságról lenne szó.

A fentiekből az is következik, hogy – a köznapi szóhasználat ellentétben – a minta reprezentativitása nem feltétlenül függ

- egy szinten túl a mintanagyságtól,
- a mintavétel véletlenszerűségétől, „vakságától”.

Azt, hogy mintánk valóban reprezentatív-e, csak az alapsokaság jellemző tulajdonságainak és a mintának az együttes ismeretében dönthetjük el.

Így a reprezentativitás olyan tulajdonság, melynek megítélése csupán a mintavételi terv ismeretében nem lehetséges. Természetesen tudatában kell lenni annak, hogy a minta csak néhány fontosabb tulajdonság, ismérv szerint lehet reprezentatív, és ezért ez az ún. képviselői jellege soha sem érheti el, csupán megközelítheti az alapsokaságot.

A fenti elvek figyelmen kívül hagyása sok félreértésnek lehet forrása. Csak néhány gyakorlati példával szeretnénk szemléltetni ezt a megállapítást:

- Tételezzük fel, hogy egy megfelelő elemszámú, véletlenül kiválasztott minta jól reprezentálja Baranya megye felnőtt lakosságát, mind területi elhelyezkedés, mind életkor, mind pedig nemek szerint reprezentatív. Amennyiben – egyéb kérdéseken túl – a MALÉV Rt. által nyújtott szolgáltatások minőségéről is tudakozódnunk kívánunk, feltehetően torz, hamis képet kapunk. Ugyanis nem kell túlságosan nagy fantázia ahhoz, hogy feltételezzük, a fenti módon reprezentatív mintában megkérdezettek nem kellő számban veszik igénybe a légitársaság szolgáltatásait, így azok megítélésről sem adhatnak kompetens képet.
- A Pécsi Tudományegyetem jelenlegi és korábban végzett hallgatói körében vizsgálhatjuk az egyetemi képzésről kialakított képet, véleményt. Amennyiben számítógépen (interneten) kapjuk meg a válaszokat, akkor az egyébként sok szempont (pl. nem, életkor, végzés éve stb.) szerint amúgy repre-

”

Reprezentatívnak tekinthetjük azt a mintát, melynek legfontosabb jellemzői nem térnek el lényegesen a sokaság megfelelő jellemzőitől.

Mindez azt is jelenti, hogy a reprezentativitás csupán viszonylagos lehet, függ a fontosnak tartott sokasági jellemzőktől.



A reprezentativitás olyan tulajdonság, melynek megítélése csupán a mintavételi terv ismeretében nem lehetséges. A minta csak néhány fontosabb tulajdonság, ismérv szerint lehet reprezentatív, és ezért ez az ún. képviselői jellege soha sem érheti el, csupán megközelítheti az alapsokaságot.



A mintavételt végző szakmai ismeretei, tapasztalatai és a megrendelő szakismerete a döntő érvek.

”

zentatív minta félreinformál. A jelenlegi hallgatók internetes hozzáférése már napi valóság, azonban a korábban végzett hallgatók közül e korszerű informatikai eszközön eljuttatott válaszok már nem reprezentatív véleményeket tükröznek. Tudjuk, hogy ma még az interneten válaszoló volt hallgatók vagy munkahelyükről, vagy otthonukból jelentkeznek, ami felfelé „torzítja” a hallgatók munkahelyi egyéni karrierjét, röviden a jobb anyagi és egzisztenciális helyzetben lévők képesek a válaszadásra elektronikus úton.

Úgy véljük: e sematikus példák is alkalmasak annak a megvilágítására, hogy egy sok szempontból reprezentatívnak ítélt minta csak akkor alkalmas megalapozott ítéletek megfogalmazására, ha az adott kérdés érzékelhető sztochasztikus kapcsolatban van azokkal a csoportképző ismervekkkel, amelyek a reprezentativitást hivatottak jelezni. Természetesen ennek eldöntésére objektív tesztek csak korlátozottan állnak rendelkezésre. Itt a mintavételt végző szakmai ismeretei, tapasztalatai és a megrendelő szakismerete lehetnek a döntő érvek.

### A mintanagyság tervezése

A statisztikai mintavétel tervezésének egyik legbonyolultabb kérdése a megfelelő nagyságú minta elemszámának meghatározása. Ismeretes, hogy a mintanagyság, a megbízhatóság és a hibahatár egymással összefüggő kategóriák. „Hüvelykujj-szabályként” kijelenthetjük, hogy azonos nagyságú minta esetén nagyobb megbízhatóságú becslés csak szélesebb intervallum (nagyobb hibahatár) mellett nyerhető. Ezen összefüggés „átrendezése” adja azt, hogy adott maximális hibahatár mellett meghatározható az a minimális mintanagyság, mellyel egy kívánt megbízhatóság elérhető.

A mintanagyság meghatározása, tervezése során kiinduló feltételként kell kezelni az alábbiakat:

- A szórás<sup>1</sup> (s) vagy relatív szórás (V) ismerete;
- Az alapsokaság elemszámának (N) és egyben címlistájának ismerete;
- Az elemzés során elvárt hibahatár ( $\Delta$ );
- Az elvárt megbízhatóság szint ( $1-\alpha$ ), amit legtöbbször a standard normális eloszlás ( $u_{1-\alpha/2}$ ) értéke képvisel.

Az alapsokaság elemszámának (N) ismerete és felhasználása egyben ún. visszatevés nélküli, egyszerű véletlen (EV) mintát eredményez, azonban ismerete hiányában is meghatározható (különösen elsődleges kalkulációként) a mintavételi terv, ami igen nagy elemszámú alapsokaság esetén nem tér el szignifikánsan az N felhasználásával készített tervtől.

A mintavételi terv készítője a sokaság elemszámát objektíven ismeri, a szórás mértékéről pedig vagy korábbi információi, vagy gyors, hozzávetőleges becslései vannak. A minta elvárható megbízhatósága általában a vizsgálat jellegétől, a megrendelő elvárásától függ. Ugyancsak a tervezés megkezdése előtt eldöntendő kérdés az elemzés során elvárt „maximális” hibahatár a priori meghatározása. A hibahatár jelzője jogosan érdemli meg az idézőjelet, mivel a mintavételi terv kialakításának szakaszában, kevés információ birtokában csupán egy-két kiemelkedően fontos paraméterre tudjuk elvárásunkat megfogalmazni. Amennyiben kellő szigorúsággal járunk el, bízhatunk abban, hogy a tényeredmények sem térnek el számottevően a kijelölt, elvárt maximális értéktől. A munka kezdetén azonban a hibahatár szubjektív meghatározása során el kell döntenie, hogy vajon a statisztikai értelemben vett megbízhatósággal dolgozunk-e, vagy ennek a számértéknek a felét tekintjük kritériumnak. A statisztikai becslések során ugyanis a pontbecsléssel meghatározott paraméterhez hozzáadva, illetve abból levonva nyerjük a konfidencia intervallumot. A megbízók azonban nem mindig vannak ennek tudatában, és ezért elvárásaikban kimondatlanul a hibahatár felét fogalmazzák

1 A szórás az elméleti, alapsokasági  $\sigma$  helyett s-sel, mintából nyert szórással helyettesítjük, mivel az esetek többségében csak mintabeli információ áll a tervező rendelkezésére a mintavételi terv elkészítése során.

2 A kiinduló összefüggések megtalálhatók pl. Hajdu Ottó - Pintér József - Rappai Gábor - Rédey Katalin: *Statisztika I.*, JPTÉ Kiadó, Pécs 1995-99. alapozó tankönyvében.

”

Adott maximális hibahatár mellett meghatározható az a minimális mintanagyság, mellyel egy kívánt megbízhatóság elérhető.



A tervezés megkezdése előtt eldöntendő kérdés az elemzés során elvárt „maximális” hibahatár a priori meghatározása. A hibahatár jelzője jogosan érdemli meg az idézőjelet.

”

meg, így is törekedve a mind pontosabb eredményre és a mind kisebb becslési intervallumra.

Legtöbbször a mintavétel tervezése során az alapsokaságra vonatkozóan még nem állnak rendelkezésre jól használható adatok, paraméterek (hiszen éppen ezek megszerzése a cél!). Ezért meg kell elégedni olyan ismert összefüggésekkel, amelyek segítséget adnak abban, hogy egy előre meghatározott maximális hibahatár és reális, vagy maximális szórás mellett minimális mintanagyságot eredményezzen.

Tekintsünk át két közismert<sup>2</sup>, gyakran használt becslést (fi-

gyelman kívül hagyva az alapsokaság elemszámát)!

1. Az alapsokasági arányra vonatkozó becslés esetén a standard hiba és a hibahatár az alábbi képletekkel határozható meg:

$$s_p = \sqrt{\frac{p(1-p)}{n}}$$

$$\Delta = u_{1-\alpha/2} s_p$$

ahol p a mintából becsült valószínűség; n a minta elemszáma.

Kihasználva azt a közismert tény, hogy  $p(1-p)$  maximuma 0,25; a megbízhatóság, a hibahatár és a mintanagyság vonatkozásában az 1. tábla állítható össze:

Megbízhatóság	Hibahatár					
	1,0%	2,0%	2,5%	3,0%	4,0%	5,0%
90,0%	6 764	1 691	1 082	752	423	271
95,0%	9 604	2 401	1 537	1 067	600	384
99,0%	16 587	4 147	2 654	1 843	1 037	663
99,9%	27 068	6 767	4 331	3 008	1 692	1 083

A minta-elemszám meghatározása az alábbi képlettel történik:

$$n = \frac{u_{1-\alpha/2}^2}{4\Delta^2}$$

2. Hasonlóan felírhatjuk mind-  
ezt alapsokasági várható ér-  
tékre vonatkozó becslésnél.  
Ez esetben a standard hiba  
és a hibahatár:

$$S_{\bar{x}} = \frac{s}{\sqrt{n}}$$

$$\Delta = u_{1-\alpha/2} S_{\bar{x}}$$

Az alapsokasági várható ér-  
tékre vonatkozóan gyakran  
nehéz elfogadható nagyság-  
rendű a priori becslést adni,  
így az elvárható hibahatár  
nagyságrendjének megállá-  
pítása is gondokba ütközik.  
Nem szünteti meg, csupán  
„enyhíti” a problémát, ha itt  
is százalékos értékben, rela-  
tív hibahatárként fogalmaz-  
zuk meg igényünket, ami  
mellett kézenfekvő, ha a szórást is ilyen mérték-  
egységben adjuk meg, relatív szórásaként (V). Ezek  
figyelembevételével átrendezhetjük a hibahatár  
fenti képletét:

$$\frac{\Delta}{\bar{x}} = \frac{V}{\sqrt{n}} \times u_{1-\alpha/2}$$

A fenti összefüggést átrendezve, megkapjuk a ki-  
választandó mintanagyságot:

$$n = \frac{u_{1-\alpha/2}^2 \times V^2}{\frac{\Delta^2}{\bar{x}^2}}$$

Így felírható az adott megbíz-  
hatósági szinthez és relatív hi-  
bahatárhoz tartozó minimális  
mintanagyság különböző - vé-  
lelmezett - alapsokasági relatív  
szórások mellett.

Az alapsokasági relatív szó-  
rás természetesen nagyon sok  
értéket felvehet, mivel inter-  
vallumáról az alábbiakat tud-  
juk:

$$0 \leq \frac{s}{\bar{x}} \leq \sqrt{N-1}$$

$$\text{ha } \bar{x} \neq 0$$

2. tábla		Relatív hibahatár					
Megbízhatóság	Relatív szórás	1,0%	2,0%	2,5%	3,0%	4,0%	5,0%
90,0%	10,0%	271	68	43	30	17	11
	25,0%	1 691	423	271	188	106	68
	50,0%	6 764	1 691	1 082	752	423	271
	100,0%	27 055	6 764	4 329	3 006	1 691	1 082
	200,0%	108 222	27 055	17 315	12 025	6 764	4 329
95,0%	10,0%	384	96	61	43	24	15
	25,0%	2 401	600	384	267	150	96
	50,0%	9 604	2 401	1 537	1 067	600	384
	100,0%	38 414	9 604	6 164	4 268	2 401	1 537
	200,0%	132 658	38 414	24 585	17 073	9 604	6 164
99,0%	10,0%	663	166	106	74	41	27
	25,0%	4 147	1 037	663	461	259	166
	50,0%	16 587	4 147	2 654	1 843	1 037	663
	100,0%	66 349	16 587	10 616	7 372	4 147	2 654
	200,0%	265 397	66 349	42 464	29 489	16 587	10 616

Számításunkban csak néhány fontosnak vélt rela-  
tív szórásértéket emeltünk ki.

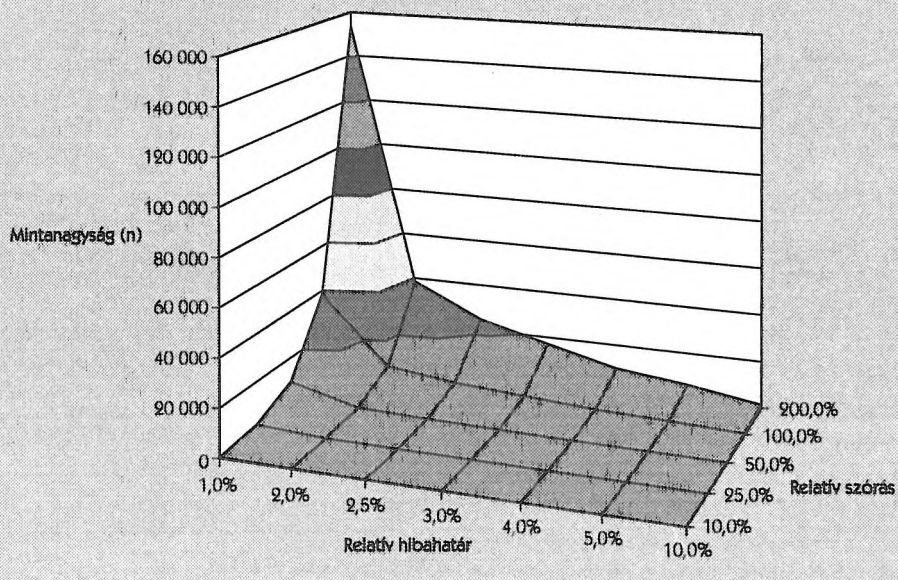
Vegyük észre, hogy ez esetben megbízhatósági  
szintenként képződik egy-egy tábla, amely a minimá-  
lis mintanagyságokat tartalmazza (2. tábla).

A 2. tábla természetét jól megismerhetjük, ha ki-  
emeljük a 95%-os megbízhatósági szintet és a kapott  
értékeket ábrázoljuk (1. ábra).

Amint az ábrából világosan leolvasható, alacsony  
relatív hibahatár erős relatív szórás mellett igen nagy

1. ábra

A minta elemszáma a relatív hibahatár és a relatív szórás függvényében  
(95%-os megbízhatóság mellett)



mintát igényel, míg a hibahatár növekedése exponenciálisan csökkenti a kívánt minta nagyságát.

Az előzőekben a mintanagyság meghatározásának kétféle módját közöltük. Összehasonlíthatjuk a kétféle megközelítést, ha feltesszük, hogy a sokasági arány (p) esetében is definiálható a relatív hibahatár ( $\Delta/p$ ). Mivel az alapsokasági arányon alapuló számítás szerint a szélsőséges  $p=0,5$  esetet használjuk fel általában, a 100%-os relatív szórás mellett találunk azonos jegyeket. (Itt kell megjegyezni, hogy míg a sokasági arányon alapuló számítás során  $p=0,5$  esetén a szórás maximumát tetelezhetjük fel, a relatív hibahatár esetén annak maximális értéke  $p=0$  közelében van.)

Az 3. tábla a kétféle megközelítés (sokasági arány, mint hibahatár illetve relatív hibahatár) mellett kapott mintanagyságokat tartalmazza 100%-os relatív szórás és 95%-os megbízhatósági szint mellett.

3. tábla							
Módszer	Hibahatár illetve relatív hibahatár						
	1,0%	2,0%	2,5%	3,0%	4,0%	5,0%	10,0%
Alapsokasági arány	9 604	2 401	1 537	1 067	600	384	96
Alapsokasági várható érték	38 414	9 604	6 146	4 268	2 401	1 537	384

A táblából kiolvashatjuk, hogy a relatív hibahatárt alkalmazva, annak fele az arányon alapuló hibahatár. A 10%-os relatív hibahatár mellett megállapított mintanagyság megfelel az 5%-os hibahatár melletti értéknek.

A mintanagyság és a hibahatár tanulmányozása érdekes összefüggésre hívja fel a figyelmet. Közismert tény, hogy a minta növekedése a hibahatár csökkenését eredményezi. Ennek mértékét és ütemét jól szemlélteti a 2. ábra.

Látható, hogy 5%-os hibahatárt 400 elemű, míg 2,5%-os hi-

”

Alacsony relatív hibahatár erős relatív szórás mellett igen nagy mintát igényel, míg a hibahatár növekedése exponenciálisan csökkenti a kívánt minta nagyságát.



Nem túl összetett vizsgálatok esetén a szokásosan alkalmazott 1000 elemű mintanagyság már kielégítő eredményeket adhat, alacsony hibahatár mellett.

”

bahatárt 1540 elemű mintanagyság alatt nem garantálhatunk.

A véges minta, illetve az egyszerű véletlenül alapuló mintavétel általában igényli az alapsokaság nagyságának (N) és elemeinek ismeretét. Ennek felhasználásával a mintanagyságot az alábbi képletek segítségével állapíthatjuk meg:

$$n = \frac{u_{1-\alpha/2}^2 \times (p(1-p))}{\frac{u_{1-\alpha/2}^2 \times (p(1-p))}{N} + \Delta^2}$$

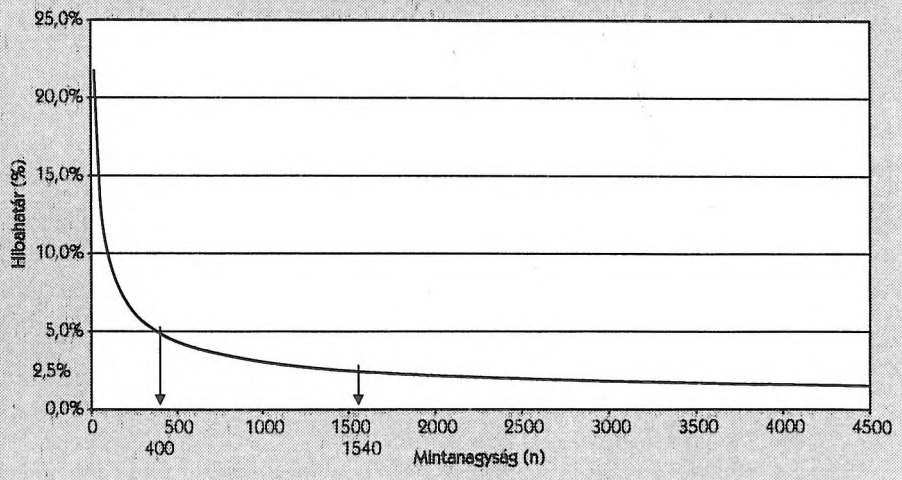
$$n = \frac{u_{1-\alpha/2}^2 \times V^2}{\frac{u_{1-\alpha/2}^2 \times V^2}{N} + \frac{\Delta^2}{\bar{x}^2}}$$

Figyelemre méltó az alapsokaság és a mintasokaság összefüggésének elemzése. A 3. ábra 5%-os hibahatár és 95%-os megbízhatóság mellett reprezentálja az összefüggést.

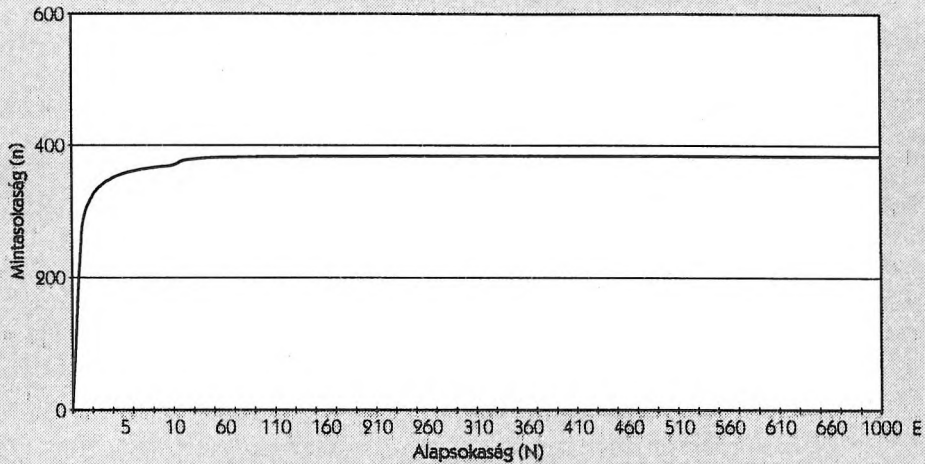
Míg egy 2000-es nagyságú alapsokaságból mintegy 320 elemű mintát kell venni, 30 000 elemű alapsokaság esetén már a mintanagyság megközelíti a maximális 384 elemet, és csak kismértékben változik. Mindez igazolja, hogy nem túl összetett vizsgálatok esetén a szokásosan alkalmazott 1000 elemű mintanagyság már kielégítő eredményeket adhat, alacsony hibahatár mellett.

2. ábra

A mintanagyság és a mintavételi hiba kapcsolata (95%-os megbízhatósági szint mellett)



3. ábra

Az alapsokaság és a mintasokaság összefüggése 5%-os hibahatárral  
(95%-os megbízhatóság mellett)

### Gyakran alkalmazott mintavételi módok

A reprezentatív adatgyűjtési módszereknek az alábbi csoportosítása az általánosan elfogadott:

#### I. Véletlen mintavétel

- Egyszerű véletlen
- Egylépcsős (csoportos)
- Többlépcsős
- Rétegzett

#### II. Nem véletlen mintavétel

- Kvóta szerinti
- Koncentrált
- Önkényes

#### III. Szisztematikus kiválasztás

Az alábbiakban elsősorban a rétegzett, és a kvóta-kiválasztás problémáira szeretnénk felhívni a figyelmet, és csupán megemlíthetjük az alap kiválasztási módokból kombinált összetettebb eljárásokat.<sup>1</sup>

A gyakorlatban – valószínűleg a végeredményben hasonló „összetételük miatt” – a rétegzett mintavételt és a kvóta szerinti kiválasztást gyakran „szinonimaként” kezelik, illetve – és ez a veszélyesebb – a kvóta szerinti kiválasztást rétegzett mintavételként interpretálják.

Nem érdektelen, ha vázlatosan összehasonlítjuk a rétegzett- és a kvóta-kiválasztás néhány jellemzőjét (4. tábla).

Lényeges különbségnek tartjuk azt, hogy a rétegzett kiválasztás során a minta összetételét mintegy

4. tábla

Rétegzett kiválasztás	Kvóta-kiválasztás
<p>A minta összetételének „javítása” növeli a minta reprezentativitását.</p> <p>Nem sérül a véletlen kiválasztás elve. Alapvetően véletlen elvű kiválasztási mód, az egyes rétegekből EV mintát veszünk.</p> <p>Amennyiben a rétegzésképző ismérv nem független a vizsgált jellemzőtől, alacsony standard hibát várhatunk.</p> <p>A minta bizonyos jellemző(k) szerinti összetétele előre tervezhető.</p> <p>A minta összetétele mindig megfelel a tervezett elvárásoknak.</p> <p>Szükséges az alapsokaság teljes lejtromának (címlistájának) ismerete.</p> <p>Az esetek többségében a becslés standard hibája kisebb lesz, mint az egyszerű kiválasztás esetében.</p>	<p>Az összeróknak bizonyos előírt arányokhoz (kvótákhoz) kell tartaniuk magukat.</p> <p>A kvóták betartása mellett az összeróknak véletlenszerűen kell mintát venniük. Amennyiben maradiéktalanul teljesül a véletlen szerepe, a minta is véletlen mintának tekinthető.</p> <p>A minta összetétele megfelel a tervezettnek, ha <math>n'</math> számú<sup>2</sup> megfigyelés az <math>n</math> elemű mintát eredményezi.</p> <p>Nincs szükség az alapsokaság lejtromára (címlistájára).</p> <p>Véletlen elvű, ha az alapsokaság elrendeződése véletlen és <math>n'</math> kiválasztás pontosan <math>n'</math> elemű mintát eredményez.</p> <p>Amennyiben a mintavétel során több a meghiusulás, a becslés hibája nagyobb is lehet, mint az egyszerű kiválasztásé.</p>

1 A reprezentatív kiválasztás, illetve statisztikai kutatás-tervezés átfogó ismertetését adja pl. Leslie Kish: Kutatások statisztikai tervezése, John Wiley & Sons, New York 1987.

2  $n'$ -vel jelöljük a kvóta-kiválasztás során nyert mintanagyságot.

„mesterségesen” javítjuk azáltal, hogy előzetesen rétegzésképző ismérveket alkalmazunk, ezáltal növeljük a minta reprezentativitását, de a kiválasztás technikája nem sérti a véletlen kiválasztás alapelvét. A minta ki-

választása az egyes szempontok szerint homogénebb csoportokból külön-külön, egymástól függetlenül történik, a csoportokként (rétegekként) egyszerű véletlen módszerrel. Így a mintába kerülés valószínűsége a rétegen belül azonos, független valószínűség. A kvótakiválasztás során az egyes megfigyelt elemek mintába kerülésének valószínűsége feltételes valószínűség, mivel függ attól, hogy az előző elem milyen réteghez, csoporthoz tartozik. Könnyen belátható, hogy míg a rétegezett kiválasztás során az elemek kiválasztási valószínűsége konstans, addig a kvótakiválasztás alkalmazása során ez nem mondható el, és a nagyobb, összetettebb minták esetében egyre kisebb annak a valószínűsége, hogy a tervezett minta jöjjön létre.

Habár mindkét mintavételi módnak sajátos jegye a tervezhetőség, a rétegezett kiválasztás esetében ez adottság. A kvótakiválasztás alkalmazásakor viszont addig gyűjtjük az adatokat (esetleg elhagyva értékes információkat!), míg a megfigyelésszám valamennyi kvótában az előírásnak (és ezáltal a tervnek) megfelelően teljesül.

A rétegzett mintavétel a legtöbb szempontból jobb valószínűségelméleti tulajdonságokkal rendelkező, becslőfüggvényeinek határeloszlásában ismert mintavétel; a kvótakiválasztásról mindez nem mondható el. Ugyanakkor nem hagyhatjuk figyelmen kívül, hogy az utóbbi a gyakorlatban lényegesen könnyebben végrehajtható, ugyanis a legkisebb elemszámú réteg kiválasztási arányszámának nem feltétlenül kell meg egyeznie a többiével. Kedvezően értékelhetjük azt a tényt is, hogy amennyiben a fenti követelményeknek megfelel a kvótakiválasztás, az eredményeket tekintve joggal sorolható a véletlenül alapuló kiválasztási módok közé. Igaz azonban, hogy a becslés standard hibáinak meghatározása még ilyen esetekben is „bonyolultabb” számításokat, kiegészítő vizsgálatokat igényel.

Röviden szólni kívánunk a gyakorlatban szintén nagyon elterjedt szisztematikus kiválasztásról. A kiválasztás jellege, megítélése az alapsokaság elemeinek elrendeződésétől függ. Amennyiben a vizsgálat céljától független a rendezési elv, a kiválasztás eredményeként kezelhető statisztikai mintát kapunk és a mintavétel rendelkezik mindazon „jó” tulajdonságokkal, mint az egyszerű véletlen mintavétel. Előnye en-

”

A kvótakiválasztás során az egyes megfigyelt elemek mintába kerülésének valószínűsége feltételes valószínűség, mivel függ attól, hogy az előző elem milyen réteghez, csoporthoz tartozik.

Míg a rétegezett kiválasztás során az elemek kiválasztási valószínűsége konstans, addig a kvótakiválasztás alkalmazása során ez nem mondható el.

”

nek a kiválasztási módnak, hogy térbeni, vagy időbeli nyilvántartás megléte nélkül is elvégezhető. Ugyancsak „kedvező” tulajdonságai között említhetjük meg, hogy amennyiben az alapsokaság egységei monoton növekvő (csökkenő) sorrendben helyezkednek el, érvényesül az ún. rétegezési effektus, tehát a nagyobb (kisebb) elemek közelebb kerülnek egymáshoz, ami önként adja a homogénebb rétegeket. Ezért ilyen esetekben joggal várhatjuk el, hogy a standard hibák kisebbek, és az elemzésben jobban felhasználhatóak lesznek. Itt is elmondhatjuk, hogy a szisztematikus mintavétel a fenti feltételek megléte esetén véletlen mintát eredményez, a

hibaszámítások elvégezhetőek.

A gyakorlatban a kiválasztási módok „vegytisztá” alkalmazására csak nagyon ritkán kerülhet sor. A célkitűzések, az elemzési igények, egyszerűbben a megrendelők kívánságai a mintavételt tervezőket összetettebb mintavételi tervek kialakítására serkentik. Így a lépcsőzetes kiválasztás gyakran párosul a rétegezett kiválasztással; gyakori igény az ún. önsúlyozó minták használata, amelyekbe beépítik egyes ismérvek fontosságának, súlyának szerepét; minden összetettebb mintavételi terv számít a szisztematikus kiválasztásra, annak gyakorlati előnyeire és nem elhanyagolható gyakorlati szempont, hogy a tényleges mintavétel sokszor a kvótakiválasztás alkalmazásával valósul meg. Mindezek a tényezők megnehezítik a becslési hibák kiszámítását, korrekt becslési intervallumok megállapítását.

\* \* \*

Tanulmányukban a mintavételi tervek készítése során óhatatlanul felmerülő problémák közül csupán néhányra szeretnénk volna felhívni a figyelmet. Valljuk azt, hogy a mintanagyság megállapítása, a mintavételi módok kiválasztása egyaránt igen jelentős hatást gyakorol a vizsgálatra, az eredmények értékelésére, interpretációjára. Tudatában vagyunk annak, hogy tökéletes mintát nem lehet elvárni, azonban törekedni kell arra, hogy az eredmények megbízhatóan, egy jól értelmezett reprezentativitás mellett kerüljenek felhasználásra.

*Pintér József egyetemi adjunktus,  
Rappai Gábor egyetemi docens, dékánhelyettes  
a PTE Közgazdaságtudományi Karán*