

Mintavételi dilemmák a könyvvizsgálatban

Dr. Ország Adrienn – Sándorné dr. Kriszt Éva

Bevezetés

A könyvvizsgáló egyik nagy problémája, **hogyan szerezzen kellő bizonyosságot** a könyvvizsgálat során úgy, hogy eltekintve az igen kis cégektől és a kis tételszámoktól, vizsgálata nem terjedhet ki minden könyvelési tételre. Elkerülhetetlen, hogy csak egy részhalmazt vizsgáljon meg, mert kapacitás-, idő- és költség- oldalról is korlátozottak a lehetőségei. Elterjedt tehát a könyvvizsgálatban is a mintavétel alkalmazása.

A könyvvizsgálatban jelenleg alkalmazott mintavételi módok

A statisztikai mintavétel kivetítési célú, a mintavételi kockázat számszerűsíthető. Véletlen mintavétel esetén a statisztikai hiba számszerűsíthető és az általánosítás megbízhatósági szintje meghatározható. A statisztikai mintavétel nemzetközi gyakorlatban is elterjedt egyik gyakori módja a pénzegység alapú mintavétel (**MUS**, azaz Monetary Unit Sampling), amelyről a későbbiekben még szólunk.

Többféle statisztikai mintavételi mód létezik, de felhívjuk a felhasználó figyelmét, hogy akármelyiket választja is, az értelmezésnél mindig ügyeljünk a **reprezentatív minta** fogalmának használatára.

A reprezentativitás, mint cél arra utal, hogy a sokaság bizonyos egységei (a minta) azzal a céllal kerülnek kiválasztásra, hogy a sokaság egységeit (sokaság egészét) megfelelő módon képviseljék. Így a minta elemzéséből **a sokaság jellemzőire következtethetünk**.

A nem statisztikai mintavétel **felderítő, elemző célú mintavétel**. Az elemző célú (nem statisztikai) mintavételeket a kockázatelemzésnél használják a könyvvizsgálatban, segítségével azon részterületeket jelöli ki a könyvvizsgáló, amelyek nagyobb eséllyel tartalmaznak hibát.

Hogyan haladjon a könyvvizsgáló a mintavétellel?

Egyáltalán milyen esetben van szükség a mintavételre és mikor lehetne attól eltekinteni?

A könyvvizsgáló a mérlegsorok és az eredménykimutatás sorai szerinti haladáson alapulva tervezi meg munkáját.

- **Első lépésben** a könyvvizsgáló kiszűri a kiugróan magas értékű tételeket és azokat egyedileg megvizsgálja. Ez tulajdonképpen megfeleltethető a koncentrált kiválasztásnak. Amennyiben ezeknek a tételeknek az összértéke lefedi a vizsgálni kívánt sokaság értékének kétharmadát, már kedvező helyzetben van a könyvvizsgáló ahhoz, hogy kellő bizonyossággal fogalmazzon meg állításokat a sokaságról.
- **Második lépésben:** kérdés, hogy a maradék halmazról hogyan nyilatkozzon. Mielőtt a mintavételi eljárások bármelyikéhez is fordulna, **érdemes tanulmányoznia az adatállományt**. Az adatállomány megismerésében segítségére lehetnek a leíró statisztikai módszerek és az ezekhez kapcsolódó mutatószámok. Ezeket általában a legegyszerűbb statisztikai programcsomagok könnyen és szinte automatikusan számítják. (Pl. az Excel-ből is könnyen átemelhetők.) Hasznos, tájékoztató

információhoz juthat a könyvvizsgáló, ha a maradék halmazból a legnagyobb és a legkisebb értékű tétel különbségét kiszámítja, ez lesz a szóródás terjedelme, amely kijelöli, hogy milyen intervallumban fordulhatnak elő értékek.

- Érdemes azonban innen továbblépni és **a tételek átlagos értékét** is meghatározni. Az átlagnak azonban rendszerint nincs tapasztalati megfelelője, hanem az átlag körül szóródnak a különböző könyvelési tételek értékei.
- Ez indokolja, hogy az átlag mögé is nézzünk és megvizsgáljuk az egyes tételeknek az átlagtól való átlagos eltérését. Ezt a **szórás** mutatja meg számunkra.
- Nem mindegy azonban, hogy mekkora átlagtól nézzük az átlagos eltérést, ezért itt nem állhatunk meg, hanem kiszámítjuk a szórás és az átlag hányadosát, amit relatív **szórásnak** hívunk és általában **százalékos** formában értelmezzük. A szokásos határ 10%, ami alatt azt mondhatjuk, hogy az átlag jól jellemzi a vizsgált értékeket, és így a sokaságunk nem mutat nagy változékonyságot.
- Sok esetben találkozhatunk olyan helyzetekkel, hogy az átlag nem jellemzi igazán jól a sokaságot, mert az értékek zöme nem az átlag körül helyezkedik el, hanem más érték körül tömörül. Az így meghatározott tömörülési helyet **módusznak** nevezzük és **tipikus** értéknek is hívjuk, mert kifejezi, hogy azokból, vagy a hozzá nagyon közel eső értékekből van a legtöbb tétel a vizsgált sokaságban. Megjegyezzük, hogy nem feltétlenül kell minden vizsgált halmazban tömörülési helynek, azaz módusznak lennie, ráadásul lehet több tömörülési hely, azaz több módusz is. Ennek feltárásában segítheti a könyvvizsgálót a vizuális megjelenítés.
- Javasoljuk grafikus ábra készítését is, mert egy kifejező ábrával, egy oszlop-, vagy vonaldiagram segítségével viszonylag gyorsan nyilatkozhatunk a könyvelési tételek érték szerinti eloszlásáról a vizsgált sokaságban. Az ábra olyan oszlopdiagram, amelynek oszlopai hézag nélkül illeszkednek egymáshoz. A vízszintes tengelyen az érték adatok szerepelnek, (célszerű, lehetőleg azonos terjedelmű -tól -ig kategóriákat képezni) a függőleges tengelyen pedig az előfordulások jelennek meg. Ha ezek után összekötjük a téglalapok felső élének középpontjait, vonaldiagramhoz jutunk, ami már szépen mutathatja a ferdeséget. (Ilyen hisztogramot és gyakorisági poligont az Excel tud készíteni.)
- Ha viszonylag jelentős ferdeség rajzolódik ki a könyvvizsgáló előtt, akkor érdemes számításokat is végezni és megnézni, hogy az **átlag és a módusz mennyivel tér el** egymástól, és amennyiben ez az **eltérés abszolút értékben meghaladja a szórás 50%-át**, akkor változékonny a sokaságunk (jelentős aszimmetria mutatkozik). Ilyenkor

érdeemes és szükséges a maradék halmazból mintát venni és további bizonyítékokat szerezni a megbízható és valós kép kialakításához az adott adatállományról.

- A leíró statisztika további jellemzőket is közöl a felhasználóval; megmutatja a **medián** és a kvartilis értékeit is. A medián az értékek nagysága szerinti sorba rendezés után a középső értéket mutatja, azaz azt az értéket, amelynél ugyanannyi kisebb és ugyanannyi nagyobb értékű tétel helyezkedik el a sokaságban.
- De nemcsak két egyenlő részre, hanem pl. négy egyenlő részre is oszthatjuk az adatállományunkat. Ebben az esetben **kvartilisek**hez jutunk. Az első kvartilis egynegyed és háromnegyed arányban osztja fel a sokaságot, míg a felső kvartilis háromnegyed - egynegyed arányban. Tehát az alsó kvartilis, a Q_1 megmutatja azt az értéket, amelynél a könyvelt tételek egynegyede kisebb értékű, háromnegyede pedig nagyobb értékű. A felső kvartilis, a Q_3 pedig az az érték lesz, amelynél a vizsgált tételek 25%-a nagyobb értékű, 75%-a pedig kisebb értékű.
- A medián és a kvartilisek közötti távolságból is következtethetünk az érték szerinti eloszlásra. Ha a medián és a felső kvartilis közötti távolság megegyezik a medián és az alsó kvartilis közötti távolsággal, akkor szimmetrikus az eloszlás, nincsenek határozott irányú eltolódások, vagyis a további mintavételtől eltekinthetünk. Amennyiben azonban ezekből az eltérésekből számított **mutató** abszolút értékben **eléri a 0,5** értéket, akkor **indokolt lehet a további mintavételes vizsgálat**. A teljesség kedvéért megjegyezzük, hogy ezt a mutatót „F” mutatónak nevezzük és az alábbi formulával számíthatjuk:

$$F = \frac{(Q_3 - M_e) - (M_e - Q_1)}{[Q_3 - M_e] + [M_e - Q_1]}$$

A formula tájékoztató jellegű, mert a statisztikai programok automatikusan megadják ennek a mutatószámnak az értékét.

Mindezek a vizsgálatok hozzájárulhatnak ahhoz, hogy a könyvvizsgáló jobban megismerje a vizsgálati sokaságot, és ha további vizsgálat, illetve mintavétel indokolt, akkor megalapozottá tegye pl. a vizsgálati lényegesség beállítását a programban.

Javaslatok a vizsgálati lényegesség értékének meghatározására:

- A számtani átlag és a tipikus érték (módusz) **különbsége** abszolút értékben **a szórás értékének kétszerese**.
- A felső kvartilis és a medián, illetve a medián és az alsó kvartilis közötti távolság (abszolút értékben) közül **a nagyobb különbség háromszorosa**.

Ha az eddigi elemzés alapján megállapítja a könyvvizsgáló, hogy mintavételes eljárást kíván alkalmazni, mert pl. az adathalmaz nagy változékonyságot mutat, és a **Pénzegységalapú**

mintavételt választja, mint nemzetközi gyakorlatban elfogadott standardot, akkor következőket kell mérlegelnie:

A pénzegység alapú mintavétel legfontosabb jellemzője, hogy **konzervatív**: a kiértékelés **érzékeny a hibák számára**. Elméleti alapját az Elemszámmal Arányos Valószínűségek teremtik meg, olyan visszatevés nélküli minta kiválasztásával, ahol minden sokasági elem mintába kerülési valószínűsége egyenesen arányos annak könyv szerinti értékével. Ez a pénzegység alapú mintavétel (**MUS**, azaz Monetary Unit Sampling) elméleti alapja.

Előnyei:

- a mintát nem szükséges rétegezni és ez egyszerűsíti a mintavételt;
- a minta elemszámának meghatározásához nem szükséges a sokaság, azaz az összetétel statisztikai tulajdonságainak (átlag, szórás, eloszlás) ismerete;
- ha a várható hiba alacsony, akkor hatékony mintaméretet ad.

Hátrányai:

- a hiba nagysága egyetlen tételnél sem haladhatja meg a könyv szerinti értéket;
- elsősorban felülértékelések felfedésére használható, az alulértékelések vizsgálatára korlátozottan alkalmas;
- amennyiben sok a várható hiba, nagyon magas mintaelemszámot eredményez.

A pénzegység alapú mintavétel lényege

A pénzegység alapú mintavétel lényege, hogy a sokaság minden egyes forintja egy mintavételi egység. Minden mintavételi egység (Ft) egy logikai egységet képvisel, tehát minden forint egy könyvelési tételhez tartozik. A sokaságot véletlenszerű sorrendbe rendezzük (bármilyen sorrend megfelelő, csupán arra kell figyelni, hogy az adatokban ne legyenek periodikusan ismétlődő tételek), majd **véletlen kezdőponttal szisztematikus kiválasztással** mintát veszünk. A szisztematikus kiválasztás azt jelenti, hogy a véletlen kezdőponttól minden „k”. forint kerül bele a mintába. A mintaelemszám meghatározása megadja a mintavételi intervallumot („k”).

A pénzegység alapú mintavétel esetén tulajdonképpen hibás forintokat keresünk, ezért akkor célszerű a használata, ha a hibás tételek kis számban fordulnak elő. Azok a tételek, amelyek nagyobbak, mint a mintavételi intervallum, biztosan bekerülnek a mintába, ezért ezeket célszerű előre kiválasztani (teljeskörű megfigyelés). A nulla Ft-os tételek ezzel szemben biztosan nem kerülnek be a mintába, ezért ezen hibák feltárására (pl. le nem könyvelt szállítói számla) más módszert kell választani. Mindezek alapján a pénzegység alapú mintavétel a felülértékelések felfedésére használható. Megállapítható, hogy amennyiben az alulértékelés 10%-a a felülértékelésnek, akkor a tévedés esélye alacsony.

A pénzegység alapú mintavétel (MUS) sajátossága, hogy egy olyan paramétert használ, mely más eljárásoknál nem fordul elő: ez a **várható hiba**. A várható hiba azt jelenti, hogy mekkora összegű hibára számítunk. Ennek a hibának a konkrét realizációja a mintából számított kivetített hiba. A **várható hiba / elfogadható hiba aránya** befolyásolja a megbízhatósági együtthatót, és ezen keresztül a minta méretét.

A minta elemszámokra a szakértői anyagokból az alábbi táblázat használatát javasoljuk:

Várható hiba az elfogadható hiba arányában	Mintavételezési kockázat	Elfogadható hiba a mintasokaság %-ában						
		30%	10%	5%	4%	3%	2%	1%
0	5%	10	30	60	75	100	150	300
0	10%	8	24	47	58	77	116	231
0	20%	6	17	33	41	54	81	161
0	35%	4	11	21	27	35	53	105
0,2	5%	16	47	93	116	155	232	463
0,2	10%	12	35	69	86	114	171	341
0,2	20%	8	23	46	58	77	115	230
0,2	35%	5	14	28	35	46	69	138

A táblázatban szerepel a mintavételezési kockázat, melynek általánosan elfogadott mértéke 5%, ugyanakkor a könyvvizsgálati gyakorlatban előfordul 20% is.

A pénzegység alapú mintavétel mindezek alapján **igen nagy sokaságok esetén hatékony**, hiszen a sokaság elemszáma nem befolyásolja a minta méretét, közvetlenül az összege sem.

Kiértékelés

A hiba felső határa = a kivetített hiba + a statisztikai mintavétel hibája

A hiba felső határát hasonlítjuk össze az elfogadható hibával, ezek alapján hozzuk meg a könyvvizsgálatról a döntést. A kivetített hiba a mintában talált hiba arányosítása a teljes sokaságra. A statisztikai mintavétel hibája pedig a mintavételi kockázat számszerűsítése. Amennyiben hibás tételt találunk, a kivetített hiba = a mintavételi intervallummal, ha csak részben hibás a tétel, akkor arányosítunk.

A megbízhatósági együtthatókra a következő táblázatban lévő értékeket javasolja a szakirodalom:

A mintában talált hibák száma	Mintavételezési kockázat			
	5 %	10 %	20 %	30 %
0	3,00	2,31	1,61	1,21
1	4,75	3,89	3,00	2,44
2	6,30	5,33	4,28	3,62
3	7,76	6,69	5,52	4,77
4	9,16	8,00	6,73	5,90
5	10,52	9,28	7,91	7,01

A következőkben bemutatunk néhány gyakorlati példát a pénzegység alapú mintavétel alkalmazására:

A megoldás során a fenti táblázatot fogjuk használni, valamint **5%-os mintavételezési kockázattal** és **800.000 Ft-os mintavételi intervallummal** számolunk.

Példák

1. Ha nem találtunk hibát

Ebben az esetben a kivetített hiba 0.

A megbízhatósági együttható 3,00, így a statisztikai mintavétel hibája $3,00 * 800.000 = 2.400.000$ Ft.

A hiba felső határa: $0 + 2.400.000 = 2.400.000$ Ft.

2. Alulértékelések

A módszer alulértékelések esetén csak igen korlátozottan alkalmazható, hiszen az alulértékelt tételek kisebb eséllyel kerülnek a mintába, mint ha ténylegesen értékelnék őket. Nem használható a módszer, ha

- van olyan hiba, amely nagyobb, mint a könyv szerinti érték;
- ha az összes hibás tétel legalább 10 százaléka alulértékelés;
- ha az alulértékelések értékösszegének abszolút értéke meghaladja a felülértékelések összegének 10 százalékát (az utóbbi két feltétel a 10%-os határ darabszámban és értékben).

Amennyiben a fentiek nem állnak fenn, akkor a hibák abszolút értékére a felülértékelések szerint kell eljárunk.

3. Felülértékelések

a.) Teljes mértékben hibás tételek, amelyek nagyobbak, mint a mintavételi intervallum

Két hibás tételt találtunk: az egyik 1 millió Ft, a másik 2 millió Ft.

A kivetített hiba (mivel mindkettő nagyobb a mintavételi intervallumnál, a 800.000 Ft-nál) = a tényleges hiba: $1 + 2 = 3$ millió Ft.

A statisztikai mintavétel hibája (ld. 1. pont): $3,00 * 800.000 = 2.400.000$ Ft.

A hiba felső határa: $3.000.000 + 2.400.000 = 5.400.000$ Ft.

b.) Teljes mértékben hibás tételek, amelyek kisebbek, mint a mintavételi intervallum

4 hibás tételt találtunk: 2 db 100 Ft-os, 1 db 5.000 Ft-os és egy 785.000 Ft-os.

A kivetített hiba (mivel mindegyik kisebb a mintavételi intervallumnál, a 800.000 Ft-nál, így a teljes intervallumot hibásnak tekintjük) = a mintavételi intervallummal: $4 * 800.000 = 3.200.000$ Ft.

A megbízhatósági együttható (fenti táblázat 4 hiba, 5%) 9,16.

A hiba felső határa: $9,16 * 800.000 = 7.328.000$ Ft.

A statisztikai mintavétel hibája: $7.328.000 - 3.200.000 = 4.328.000$ Ft.

c.) Részben hibás tételek, amelyek nagyobbak, mint a mintavételi intervallum

2 részben hibás tételt találtunk: az egyik könyv szerinti értéke 1.000.000 Ft, tényleges értéke 900.000 Ft (hiba 100.000 Ft), a másik könyv szerinti értéke 1.200.000 Ft, tényleges értéke 700.000 Ft (hiba 500.000 Ft).

A kivetített hiba a hibák értékének összege: $100.000 + 500.000 = 600.000$ Ft.

A megbízhatósági együttható 3,00, így a statisztikai mintavétel hibája $3,00 * 800.000 = 2.400.000$ Ft. Ennek oka, hogy a talált hibákhoz tartozó logikai egységek lefedik a mintavételi intervallumot (ld. 1. pont).

A hiba felső határa: $600.000 + 2.400.000 = 3.000.000$ Ft.

d.) Részben hibás tételek, amelyek kisebbek, mint a mintavételi intervallum

5 részben hibás tételt találtunk:

Sorszám	Könyv szerinti érték	Tényleges érték	Hiba
1.	1.000	100	900
2.	500.000	400.000	100.000
3.	10.000	9.000	1.000
4.	200.000	20.000	180.000
5.	250.000	120.000	130.000
Összesen	961.000	549.100	411.900

Megnézzük a hibák százalékos arányát, és a kivetített hiba a mintavételi intervallum arányos részével lesz egyenlő:

Sorszám	Könyv szerinti érték	Hiba	Hiba aránya (%)	Kivetített hiba
1.	1.000	900	90,00	720.000
2.	500.000	100.000	20,00	160.000
3.	10.000	1.000	10,00	80.000
4.	200.000	180.000	90,00	720.000
5.	250.000	130.000	52,00	416.000
Összesen	961.000	411.900	42,86	2.096.000

A táblázatból jól látható, hogy az 1. tételnél talált 900 Ft-os hiba jelentősen nagyjítja a kivetített hibát, ezért a nagyon kis értékű tételeket célszerű előre kivenni.

A kivetített hibákat csökkenő sorrendbe rendezzük, majd a megbízhatósági együttható változását a kivetített hibákhoz rendeljük:

Sorszám	Kivetített hiba	Megbízhatósági együttható	Megbízhatósági együttható változása	Hiba felső határa
	.	3,00	-	
1.	720.000	4,75	1,75	1.260.000
4.	720.000	6,30	1,55	1.116.000
5.	416.000	7,76	1,46	607.360
2.	160.000	9,16	1,40	224.000
3.	80.000	10,52	1,36	108.800
Összesen	2.096.000	-	-	3.316.160

A statisztikai mintavétel hibája: $3.316.160 - 2.096.000 = 1.220.160$ Ft.

Összegezve a felülértékelések példáit:

Eset	Tényleges hiba	Kivetített hiba	Statisztikai mintavétel hibája	Hiba felső határa
a.)	3.000.000	3.000.000	2.400.000	5.400.000
b.)	790.200	3.200.000	4.128.000	7.328.000
c.)	600.000	600.000	2.400.000	3.000.000
d.)	549.100	2.096.000	1.220.160	3.316.160
Összesen	4.939.300	8.896.000	7.748.160*	16.644.160

*Mivel a 0 hibás megbízhatósággal kétszer számoltunk, ezért az egyiket le kell vonni (2.400.000).

A hiba felső határát végül összehasonlítjuk az elfogadható hibával, ami alapján meghozzuk a döntést.

Következtetések, javaslatok

- A mintavétel alkalmazása a könyvvizsgálatban elkerülhetetlen, de bármelyik mintavételi módot választjuk is, az értelmezésnél mindig ügyelni kell a reprezentatív minta fogalmának használatára.
- A reprezentativitás, mint cél arra utal, hogy a sokaság bizonyos egységei (a minta) azzal a céllal kerülnek kiválasztásra, hogy a sokaság egységeit (sokaság egészét) megfelelő módon képviseljék. Így a minta elemzéséből a sokaság jellemzőire következtethetünk.
- Ne használjuk a reprezentatív minta kifejezést a mintavételből származó adatok jellemzésére anélkül, hogy pontosan leírnánk, miként biztosítottuk, hogyan ellenőriztük azt.
- A reprezentativitásra (bármely értelmezésben) nem csak a minta választásakor, hanem az adatfelvételi folyamat minden lépésében oda kell figyelniünk, pl. már a mintavételi keret megválasztásakor is. Lásd: Magyar Tudományos Akadémia SJTB Statisztikai Albizottság ajánlása: (2020. január 21.). Indokolás és további részletek elérhetők a STAB honlapján: http://www.ksh.hu/mta_sjtb.