

BOROS PÉTER–MEDVEGYEV PÉTER

Közvélemény-kutatások statisztikai szemszögből – ahol biztosan hibáztak

Jelen munka a brexitet és az amerikai elnökválasztást megelőző közvélemény-kutatások hibás előrejelzéseivel és a hibák okaival foglalkozik. Megvizsgáljuk az elemi statisztikai módszer, a populációarány-becslés szakirodalmát, és rámutatunk arra, hogy a közvélemény-kutatások során felhasznált minta nagysága elmaradt a pontos becsléshez szükséges szinttől. Bemutatjuk, hogy egy ilyen jellegű becslés technikája mögött meghúzódó normális eloszlás feltételezése nem minden esetben helytálló, így következtetéseink a statisztikai módszerek széles körét érintik.*

Journal of Economic Literature (JEL) kód: C83, D72.

„Komoly brexit-hátrányt jósol a legújabb felmérés. [...] A Populus közvélemény-kutató szerint 55–45 százalék lehet az eredmény a maradáspártiak javára.” (Zubor [2016])
„A többség most bent maradna. [...] A The Mail on Sunday című vasárnapi konzervatív brit napilap megbízásából a Survation közvélemény-kutató által készített vizsgálat 45:42 százalékos megoszlást mutatott ki a bennmaradás híveinek javára.” (D. Sz. [2016])

A fenti címsorok mindenkinek ismerősek lehetnek, hiszen a 2016-os nyár folyamán a brexit-népszavazásról szóló híreket lehetetlen volt elkerülni. Ez cseppet sem meglepő, hiszen talán nem túlzás azt állítani, hogy az új évezred egyik legnagyobb társadalmi és politikai változása várt ránk. Ma már mindenki tudja azt is, hogy az esemény fontossága mellett hasonló méretűre nőtt a végső kimenetet övező meglepetés is. A britek döntése, miszerint kilépnének az Európai Unióból, rávilágított egy addig rejtett problémára: az ilyen horderejű döntéseket kísérő bizonytalanság csökkenésében legfontosabb feladatot ellátó intézmények, a közvélemény-kutatók hibásan

* A tanulmányban kifejtett nézetek kizárólag a szerzők személyes véleményét tükrözik.

Medvegyev Péter, Budapesti Corvinus Egyetem, Matematika Tanszék (e-mail: medvegyev@uni-corvinus.hu).

Boros Péter, Budapesti Corvinus Egyetem, PhD-hallgató (e-mail: borospeter90@gmail.com).

A kézirat első változata 2016. november 23-án érkezett szerkesztőségünkbe.

DOI: <http://dx.doi.org/10.18414/KSZ.2017.12.1265>

működnek. A választás előtti utolsó tíz napban nyilvánosságra hozott eredmények alapján a közvélemény-kutatások nagy többsége a maradáspártiakat hozta ki győztesül.¹ Jelen tanulmányban megpróbálunk magyarázatot adni arra, hogy mi is okozhatta a közvélemény-kutatások sikertelenségét.

A téma fontosságának szemléltetéséhez meg kell értenünk a közvélemény-kutatások általános szerepét, amit három pontban foglalunk össze. Elsőként, a kampányoló felek, így például választásokkor az elnökjelöltek vagy a politikai pártok a közvélemény-kutatások eredményeiből képesek következtetéseket levonni a szavazás végkimeneteléről. A mintából megismert pillanatnyi helyzet alapján mérik fel, hogy szükségük van-e a pozíciójuk javítására és annak megfelelően stratégiájuk alakítására vagy épp további erőforrások bevonására, ami a becslések költség-optimalizálásban betöltött szerepét emeli ki. Az ilyen jellegű felmérések második kulcsfeladata a választók tájékoztatása és a társadalom, a piacok vagy a nemzetközi szervek felkészítése az adott kimenetre. Ez különösen fontos a média felerősítő szerepe mellett. A média a közvélemény-kutatások „tőkeáttételeként” működik, hiszen gyakran felnagyítja az eredmény relevanciáját, miközben a részleteket eltörpíti. Végül a közvélemény-kutatásoknak befolyásoló hatásuk van. A nemzetközi szakirodalom két fő elméletet különböztet meg: a „tarts a győztesrel” (*bandwagon*) és a „segítsük a gyengét” (*underdog*) hatást. Amikor egy választó inkább a már vezető oldalra szavaz a közzétett eredmények befolyása miatt, akkor a *bandwagon* határról beszélünk. Ellenkező esetben, amikor inkább a vesztesre álló felet támogatják, hogy annak hátránya csökkenjen, az *underdog* elmélet teljesül. A brexit során is úgy tűnik, hogy az idő múlásával változtak a közvélemény-kutatások eredményei. Így nem zárhatjuk ki ezek befolyásoló hatását sem.²

A brexitszavazás utáni hónapokban az angol font történelmi mélypontra zuhant a dollárral szemben, a részvénytőzsi indexek szinte kivétel nélkül csökkentek, és máig tartó bizonytalanság került a pénzügyi piacokra. A brexit pontos következményeit sem a szavazás előtt, sem pedig ma nem lehet meghatározni, de azt feltételezhetjük, hogy ha a szavazást megelőző időszakban mindenki a kilépésre készült volna, akkor az azzal járó költségek sokkal enyhébbek lehettek volna. A köztudatban azonban a közvélemény-kutatások félrevezető eredményei terjedtek el, amelyek a maradást tették meg esélyesebb kimenetelként.

Nem kellett sokat várni, az amerikai elnökválasztás során újra csődöt mondtak a közvélemény-kutatók. Jellemzően az összes felmérés Hillary Clintont hozta ki győztesül, ami alapján adódik a kérdés: hogyan lehet az, hogy a közvélemény-kutató intézmények szinte következetesen rossz eredményeket közöltek?

¹ A kilépéspártiak előnyét ebben az időszakban három intézmény mérte: az Opinium, a TNS és a YouGov. Az érdemüket csökkenti, hogy eredményükben nem korrigáltak a bizonytalanokkal, akik a három esetben a teljes minta 9–16 százalékát adták. Másrészt, hogy ugyanebben a tíznapos intervallumban a YouGov kétszer mért maradáspárti előnyt, míg az Opinium egyszer döntetlent. Az összes intézményt figyelembe véve, egyetlen olyan felmérés sem volt, amely a bizonytalan válaszok utáni korrigálást követően a kilépést hozta volna ki eredményül.

² A választás után számos blogbejegyzés taglalta a brexit előjelzésének kudarcát (lásd például Fry [2016] a befolyásoló hatásokat is tesztelő blogbejegyzését).

Jelen munkában a közvélemény-kutatások módszertanával foglalkozunk. Egy társadalmi választás előrejelzése különösen nehéz feladat. A pontos eredményhez alaposan megtervezett felmérés szükséges, amely egy bonyolult optimalizációs problémaként is felfogható. Számos faktor, mint például a mintavétel hibái, a bizonytalanok száma, a reprezentativitás hiánya vagy éppen a preferenciák elrejtése, illetve változása torzíthatja a mintából adható előrejelzést. A tervezés során azonban az egyik legfontosabb formalizálható kérdés a mintanagyságra vonatkozik: mekkora mintára is van szükség ahhoz, hogy a közölni kívánt eredmény közel álljon a valódihoz? Ez a kérdés motiválja jelen munkánkat, ugyanis véleményünk szerint a brexitet és az amerikai elnökválasztást is előre jelző intézmények itt hibázták a legnagyobbat. Tanulmányunk a minta elemszámával foglalkozik, a fő állításunk azonban nem az, hogy a hibás közvélemény-kutatások egyetlen oka a rosszul megválasztott mintanagyság volt. A legutóbbi közvélemény-kutatások kudarcainak egy adott szemszögből történő elemzése mellett írásunk célja, hogy felhívja a figyelmet a közgazdasági becslések során használt mintaelemszám megállapításának fontosságára.

A következő részekben a statisztika egyik legalapvetőbb kérdésével, a populációarány becslésével foglalkozunk. A témához tartozó szakirodalom alapján megvizsgáljuk, hogy a mintavétel mely jellemzőit kell figyelembe venni a döntéskor. Ugyan a munkánk a brexitre összpontosul, az általunk levont következtetések a statisztikai becslések széles körét érintik, és így azokkal szembeni óvatosságra is intenek.

Mekkora is a minta?

A brexit vagy éppen az amerikai elnökválasztás előtti közvélemény-kutatásokat felfoghatjuk úgy is, mintha egy arány megbecslését tűzték volna ki célul: az Egyesült Királyság lakosságának mekkora aránya is szeretne kilépni az Európai Unióból; vagy: mekkora Donald Trump támogatottsága a jelenlegi állás szerint?³ Az ilyen jellegű becslések a statisztika elemi módszerei közé tartoznak, és ezek bemutatása minden bevezető statisztikakurzus része. Tehát azt gondolhatnánk, hogy a tudomány már részletesen kidolgozta a becslés módszertanát, és így képesek vagyunk pontos becslést adni. Ez részben igaz, de a helyzet korántsem ilyen egyszerű.

Egy populációs arányszám megbecslését három mennyiséggel jellemezhetjük: a minta eredménye, a hibahatár és a konfidenciaszint. Ennek megfelelően egy közvélemény-kutatás x eredményt mutat d hiba mellett $(1 - \alpha)$ százalékos konfidenciaszinten.

A hibahatárt definiálhatjuk úgy, mint a tényleges érték és a becslés eltéréseinek az esetek jelentős részében vett maximális értéke.⁴ Tehát ha a szavazás valódi értéke 50 százalék valamelyik kimenetre, akkor a valódi érték valahol $(50 - d)$ százalék és $50 + d$ százalék között van bizonyos valószínűséggel. A lefedett valószínűséget a

³ Az amerikai elnökválasztáson más jelöltek is indultak. Lényegében azonban a választás a demokráta és a republikánus párt jelöltjei közötti harcra összpontosul, így nem hibázunk nagyot, ha itt is két lehetséges kimenetet tételezünk fel.

⁴ Köszönjük az egyik bírálónak a fontos pontosítási javaslatokat. A tanulmányt több helyen is az ő tanácsai alapján változtattuk meg.

szignifikanciaszint jelenti. Így például 95 százalék megbízhatóság mellett 20 esetből 19 esetben a hiba a fenti intervallumban lesz.

A közvélemény-kutatóknak számos kérdésben dönteniük kell a felmérés megtervezésekor. Fontos, hogy meghatározzák a megkérdezettek körét, hiszen a mintának reprezentatívnak kell lennie, azaz a társadalom álláspontját kell tükröznie. Számos intézet egy bizonyos mértékben nyilvánosságra is hozta a módszertanát, részletesen kifejtve, hogy milyen úton éri el a válaszadókat, legyen az telefonos vagy esetleg személyes kapcsolatfelvétel. Ezek a módszerek az évek során a figyelem középpontjába kerültek. Így például jó néhány tanulmány részletezi, hogy a véletlen tárcsázásra mi az optimális eljárás, vagy hogy bízhatunk-e az online leadott válaszokban. Hasonló témában találhatunk újságcikkeket, amelyek arra a következtetésre jutottak, hogy az emberek nem vállalják valódi preferenciájukat például Trump esetében (*Edsall* [2016]). Ezek valóban befolyásoló tényezők, de egy közvélemény-kutatás tervezésekor egy elemibb kérdésnek kell az elsők között felmerülnie: mekkora mintára van szükség? A brexitet jellemzően kétezer, ezer vagy éppen kevesebb elemszámú mintával próbálták előjelezni. Tanulmányunkban azt a kérdést vizsgáljuk, hogy egy több tíz vagy száz milliós ország lakói vélekedésének megismeréséhez elég lehet-e egy kétezer fős minta. Első lépésként ezt elemezzük, felhasználva az elemi statisztika eszközeit.

Számos mintanagyságról szóló munka *Cochran* [1977] könyvére hivatkozik. Vizsgáljuk meg a társadalmi arány becslésének e könyvben adott elméleti keretrendszerét. Legyen P a becsülni kívánt társadalmi arány, és tegyük fel, hogy egy véletlen mintából p becslést adunk. Feltételezzük, hogy p normális eloszlású.⁵ *Cochran* [1977] szerint a feltétel teljesülését a mintanagyság (n) megadásakor ellenőrizhetjük. A becslés célja, hogy az (1) egyenlet teljesüljön:

$$\mathbb{P}(|p - P| \geq d) = \alpha, \quad (1)$$

ahol \mathbb{P} a valószínűségi mérték, d és α a már bevezetett hibahatár és szignifikanciaszint. Azaz az esetek nagy többségében a becslült érték a valódi értéktől kicsivel tér el:

$$\mathbb{P}(-d \leq p - P \leq d) = 1 - \alpha. \quad (2)$$

Cochran [1977] levezeti, hogy a becslés szórása N számú teljes populációt és n számú mintát feltételezve a következő lesz:

$$\sigma_p = \sqrt{\frac{N-n}{N-1}} \sqrt{\frac{P(1-P)}{n}}. \quad (3)$$

Mivel feltételeztük p normális eloszlását, így standardizálás után a következő teljesül:

$$\mathbb{P}\left(-z_{\alpha/2} \leq \frac{p-P}{\sigma_p} \leq z_{\alpha/2}\right) = 1 - \alpha, \quad (4)$$

ahol $z_{\alpha/2}$ a standard normális eloszlásból származó kritikus érték.

⁵ A normalitás feltételezésekor *Cochran* [1977]-et követtük. Mivel p egy a $[0, 1]$ intervallumbeli mennyiség, így ez a feltételezés nem tökéletes. Hamarosan megmutatjuk azonban, hogy ez a lépés nem is szükséges.

Számos tankönyv szerint ugyanehhez az eredményhez juthatunk, ha közvetlenül nem tesszük fel p normalitását. Tegyük fel, hogy két lehetséges kimenet van. Véletlen minta esetén, ha X_i az i -edik válaszadó voksát jelenti, akkor az X_i Bernoulli-féle valószínűségi változó lesz. A mintából számolt arány előáll, ha a megkérdezettek közül egy adott kimenetre voksolók számát elosztjuk a minta számával, azaz:

$$p = \frac{\sum_{i=1}^n X_i}{n}. \tag{5}$$

Két lehetséges kimenet esetén az $X = \sum_{i=1}^n X_i$ egy binomiális valószínűségi változó. Ezen a ponton a statisztikatankönyvek azt állítják, hogy a (4) összefüggés ugyanúgy teljesülni fog a centrális határeloszlás tétele miatt. A tétel szerint, ha (Y_1, Y_2, \dots, Y_n) független, azonos eloszlású valószínűségi változók sorozata, amelyekre $E(Y_1) = \mu$, $Var(Y_1) = \sigma^2$ véges, és

$$Y = \frac{\sum_{i=1}^n Y_i}{n} - \mu}{\frac{\sigma}{\sqrt{n}}}, \tag{6}$$

akkor Y aszimptotikusan tart egy standard normális eloszlású valószínűségi változóhoz.

Fogadjuk el a normális eloszlás feltételezését, és kövessük Cochran [1977] munkáját! Például $1 - \alpha = 95$ százalék valószínűség mellett a relatív hibára a következő összefüggést adhatjuk:

$$d = z_{0,025} \sigma_p = 1,96 \sqrt{\frac{N-n}{N-1}} \sqrt{\frac{P(1-P)}{n}}. \tag{7}$$

Ebből visszszámolva n értékét, megkapjuk a mintaszám meghatározásához szükséges formulát:

$$n = \frac{z_{\alpha/2}^2 P(1-P)}{d^2} \left/ \left\{ 1 + \frac{1}{N} \left[\frac{z_{\alpha/2}^2 P(1-P)}{d^2} - 1 \right] \right\} \right. \tag{8}$$

Vegyük észre, hogy ha N nagy – ahogy azt feltételezhetjük mind a brexit, mind pedig az amerikai elnökválasztások esetében –, akkor a nevezőben szereplő második tag nulla közeli lesz, és így kiigazító hatása eltörpül. Másrészt azonban a számlálóban ott marad a becsülni kívánt P érték. Cochran [1977] a közelítéshez P helyett annak becsülését (p) használja. Ekkor a (8) egyenlet következő alakra redukálódik:⁶

$$n_0 = \frac{z_{\alpha/2}^2 p(1-p)}{d^2}. \tag{9}$$

⁶ Ha N nem elég nagy, akkor n_0 alapján megbecsülhetjük a tényleges minta nagyságot: $n = \frac{Nn_0}{N-1+n_0}$. Egy másik lehetőség, hogy a relatív hibát rögzítjük: $P \left(\frac{|NP - NP|}{NP} \geq r \right) = \alpha$. Ebben az esetben a mintaszámot a következőképpen adhatjuk meg: $n_0 = \frac{z_{\alpha/2}^2 p}{r^2 q}$.

A (2) és a (9) egyenlet alapján megadhatjuk a konfidenciaintervallum alsó és felső határát az alábbi formában:

$$p_0 = p \pm z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}. \quad (10)$$

Ez az eredmény a szakirodalomban Wald-módszer néven vált ismertté. Az így becsült konfidenciaintervallumot Wald-féle konfidenciaintervallumnak nevezzük, mert ez a Wald-féle hipotézisvizsgálat eredményének a megfordítása. A $H_0: p = p_0$,

$H_1: p \neq p_0$ hipotézis ugyanis a $z = \frac{\hat{p} - p_0}{\sqrt{\hat{p}(1-\hat{p})/n}}$ tesztstatisztikát használja (Agresti-Coull [1998]).⁷

A (9) egyenlet tehát megadja a minta nagyságának minimumát, amely mellett a kapott eredmény d hibát fog mutatni a $z_{\alpha/2}$ értéknek megfelelő konfidenciaszinten. A (10) egyenlet alapján már azt is láthatjuk, hogy a konfidenciaintervallum hossza ebben az esetben éppen $2d$. Tehát a közvélemény-kutatás megkezdése előtt meg kell határozni, hogy mekkora hibahatárt is szeretnénk látni. A brexit és az amerikai elnökválasztás esetében a több hónapos kampányidőszakot folyamatos közvélemény-kutatások kísérték. Így a kezdeti felmérések után már mindenki előtt világossá vált, hogy mindkét döntés esetében nagyon szoros, 50 százalék közeli mintaarányal kell számolni. Ezért a közvélemény-kutatást folytató intézményeknek a d értéket nagyon alacsonyra, ideális esetben 1 százalék alá kellett volna beállítaniuk. Ahogy azt hamarosan megmutatjuk, a valóságban nem ez történt, hiszen a hibahatár minden esetben sokkal magasabb volt. Ez alapján felmerül, hogy a közvélemény-kutatók egyfajta standardizálást használnak a d hibanagyság meghatározásakor. A lehetséges standardizálás forrását a szakirodalomban kereshetjük.

Bartlett és szerzőtársai [2001] mintanagyságról szóló munkájukban – Krejcie-Morgan [1970] munkájára hivatkozva – azt írják, hogy kategorikus változókra 5 százalék, míg folytonos változókra 3 százalék az általánosan elfogadott érték. Asher [2011] szerint egy 4 százalékos mintavételi hiba normálisnak nevezhető, de megemlíti, hogy ha a kimenet 50 százalék közelében van, akkor ezzel vigyázni kell. Cochran [1977] a hibaérték meghatározásáról röviden, de lényegretörően fogalmaz: „Ezt az értéket a mintából származó eredmény felhasználásának fényében, legjobb tudásunk szerint kell meghatározni.” (74. o.) Megemlíti, hogy az értékről leggyakrabban véletlenszerűen döntenek. Rea-Parker [2014] egy sokkal frissebb, átfogó útmutató a közvélemény-kutatások szakértőinek. A mintavételezési hiba meghatározásának folyamatáról a következőt mondja: „A döntés meghozatalakor nincsenek rögzített kritériumok. A kutatónak esetről esetre kell választania a felmérés célkitűzésével összhangban.” (165. o.) Végül ők is megjegyzik, hogy kategorikus változókra a 3 százalék vagy 5 százalék általában kielégítő. Összességében tehát azt látjuk, hogy néhány tankönyv egyfajta alapértéknek konkrétan 3 és 5 százalék közötti értéket javasol – miközben ködösen óvatosságra intenek –, s a kategorikus változó mértéke sehol sincs specifikálva. Ezek alapján azt gondolhatnánk, hogy

⁷ Ahogy az Agresti-Coull [1998] munkájában is olvasható, az így kapott konfidenciaintervallum az egyik legrégebbinek számít, hiszen már Laplace [1812] könyvében is feltűnik. Cochran [1977] munkáját tehát nem azért ismertetjük, mert azt feltételezzük, hogy ő dolgozta ki a fentebb leírt aránybecslést, hanem mert számos mű is ezt a szakirodalmat használja.

szoros verseny esetén elvárható lenne a hibahatár szintjének 1 százalékos alá csökkentése. Döntetlen körüli helyzetben ugyanis egy 3 százalékos széles konfidenciaintervallum nem elégséges annak eldöntésére, hogy melyik fél is van előrébb.

Az 1. táblázatban szemléltetjük a (9) egyenletből számolt szükséges mintaszámokat különböző feltételezések mellett. Az általunk vizsgált valós esetekben 50 százalékos körüli társadalmiarány-becslést kaptunk, így $p(1-p)$ helyére annak maximális értékét (0,25) helyettesítjük. Így a (9) egyenlet alapján adott konfidenciaszint mellett egyértelmű meghatározás adódik d és n között. Egyrészt meglepően alacsony mintaszámokat kapunk 3 és 5 százalékos hiba mellett. Másrészt azonban azt látjuk, hogy 1 százalékos vagy az alatti hibához jelentősen nagyobb mintára van szükség. Például 95 százalékos konfidenciaszinten több mint 9600 megfigyelésre van szükség ahhoz, hogy a konfidenciaintervallum 2 százalékos széles legyen.

1. táblázat

Szükséges mintaelemszám

| Konfidencia- szint (százalék) | Hibahatár (d) | | | | | | | | | |
|-------------------------------------|-------------------|------|------|------|------|------|------|------|--------|--------|
| | 5 | 4,50 | 4 | 3,50 | 3 | 2,50 | 2 | 1,50 | 1 | 0,50 |
| | százalék | | | | | | | | | |
| 90 | 271 | 334 | 423 | 552 | 752 | 1082 | 1691 | 3006 | 6 764 | 27 055 |
| 95 | 384 | 474 | 600 | 784 | 1067 | 1537 | 2401 | 4268 | 9 604 | 38 415 |
| 99 | 663 | 819 | 1037 | 1354 | 1843 | 2654 | 4147 | 7372 | 16 587 | 66 349 |

2. táblázat

Tíz nappal a brexitszavazás előtti közvélemény-kutatások

| Intézmény | Maradás- pártiak | Kilépés- pártiak | Nem döntött | Minta- elemszám | d | |
|--------------------|---------------------|---------------------|----------------|--------------------|-------------|-------------|
| | | | | | 95 százalék | 99 százalék |
| | százalékaránya | | | | | |
| Populus | 55 | 45 | n. a. | 4700 | 1,43 | 1,88 |
| YouGov | 51 | 49 | n. a. | 3766 | 1,60 | 2,10 |
| Ipsos MORI | 49 | 46 | 1 | 1592 | 2,46 | 3,23 |
| Opinium | 44 | 45 | 9 | 3011 | 1,79 | 2,35 |
| ComRes | 48 | 42 | 11 | 1032 | 3,05 | 4,02 |
| TNS | 41 | 43 | 16 | 2320 | 2,03 | 2,68 |
| Survation/IG Group | 45 | 44 | 11 | 1003 | 3,09 | 4,07 |
| YouGov | 42 | 44 | 13 | 1652 | 2,41 | 3,17 |
| ORB/Telegraph | 53 | 46 | 2 | 800 | 3,46 | 4,56 |
| Survation | 45 | 42 | 13 | 1004 | 3,09 | 4,07 |
| YouGov | 44 | 43 | 9 | 1694 | 2,38 | 3,13 |
| Opinium | 44 | 44 | 12 | 2006 | 2,19 | 2,88 |

A 2. táblázatban a brexitszavazás előtti tíz napban végzett közvélemény-kutatások eredményeit mutatjuk. Látható, hogy a legutolsó felmérések mintanagysága nagyobb volt, mint a hetekig jellemző egy-két ezer fő, de vegyük észre, hogy a mintaelemszámok így is elmaradnak az 1. táblázatban alacsony hiba mellett elvártaktól. A (9) képlet segítségével visszaszámoltuk a mintavételezési hibát 95 és 99 százalék szignifikanciaszintet feltételezve. A 2. táblázat jól mutatja, hogy miután hetekig 50 százalék körüli eredményeket mértek, még a legutolsó közvélemény-kutatások sem vették figyelembe a hibahatár megfelelő szintre történő beállítását.

A Wald-féle konfidenciaintervallum hibái

A brexitet megelőző közvélemény-kutatások rendre az általunk szükségesnek tartott mintanagyság töredékét használták. Mivel az eltéréseket az előző részben leírt becslési módszer egyszerűsége is okozhatja, ezért érdemes újragondolni a fenti levezetést. Először is azzal a feltételezéssel éltünk, hogy a mintából számolt p -érték normális eloszlást követ, avagy a centrális határeloszlás tételében kimondott konvergencia már alacsony mintanagyság mellett is teljesül. Ebből adódóan a (9) képletben a z -érték a szignifikanciaszintnek megfelelő normális eloszlás kvantilisét jelenti. Ennél a pontnál érdemes megvizsgálni, hogy mekkora elemszámra van szükség a centrális határeloszlás tételének alkalmazásához.

A bevezető statisztikatankönyvekben először a mintaátlag becslése résznél merülnek fel a szignifikanciaszintek. Számos tankönyv például a harminc elemből álló mintát már nagy mintának tekinti (Wackerly és szerzőtársai [2008] 339. o., Weiss [2017] 343. o. vagy Heumann–Schomaker [2017] 217. o.). Mások a mintaaránybecslésénél az $np(1-p) > 5$ vagy együtt az $n(1-p) > 10$ és $np > 10$ feltételeket várják el a normális közelítéshez (például Rosner [2015] 189. o. vagy Meeker és szerzőtársai [1991] 145. o.).⁸ Devore [2015] megjegyzi, hogy az alapeloszlás is számít. Véleménye szerint bizonyos eloszlásokra akár 40-50 elemből álló minta sem lehet elég, míg például egyenletes eloszlás esetén akár 12 elemre is jó közelítést ad a centrális határeloszlás tétele. Ennek ellenére hüvelykujjszabályként ő is az $n > 30$ küszöböt adja meg a tétel alkalmazásának feltételeként (226. o.). Mivel a konvergencia sebessége nem része a centrális határeloszlás tételének, így az önmagában nem elég annak eldöntésére, hogy harminc elem elegendő-e a jó közelítéshez. A konvergencia sebességére analitikus formula nem adható, de felső becslésként a Berry–Esseen-tételt alkalmazhatjuk, amely a következőt állítja.

Legyen Y_1, Y_2, \dots, Y_n független, azonos eloszlású minta, ahol $E(Y_i) = 0$, $E(Y_i^2) = \sigma^2$ és $E(|Y_i|^3) = \rho$. Ha $F_n(x)$ jelöli a $\frac{\sum Y_i/n}{\sigma/\sqrt{n}}$ eloszlásfüggvényét az x helyen, akkor

$$|F_n - \Phi(x)| < \frac{C\rho}{\sigma^3\sqrt{n}}, \quad (11)$$

⁸ Ez a feltétel a legalább harmincelemű minta feltételhez hasonló, de p -re érzékeny megkötés.

ahol $\Phi(x)$ a standard normális eloszlásfüggvény az x helyen, és C egy konstans. Az elmúlt évtizedekben C értékét újabb és újabb eredmények szignifikánsan csökkentették, és a legutóbbi becslések már 0,4748 alá teszik (*Shevtsova* [2014]).

A 3. tábla a Berry–Esseen-tételben szereplő (11) egyenlet jobb oldala alapján számolt felső határt mutatja egyenletes eloszlású mintára. Ez alapján például az egyenletes eloszlású minta átlagának eloszlásfüggvénye jelentősen eltér a normális eloszlásától még néhány száz elemű minta esetén is. Tehát a Berry–Esseen-tétel alapján még a jól közelítő eloszlásra példaként gyakran használt egyenletes eloszlás is távol áll a normális eloszlástól. Ez tehát nem visz közelebb annak megértéséhez, hogy a statisztikatankönyvek miért éppen harminc megfigyelésnél húzzák meg a kis- és a nagyminta közötti határvonalat.

3. táblázat

A Berry–Esseen-tétel alapján számolt felső határ egyenletes eloszlású mintára

| n | Különbség (százalék) | n | Különbség (százalék) | n | Különbség (százalék) |
|-----|----------------------|-----|----------------------|------------|----------------------|
| 30 | 11,26 | 200 | 4,36 | 1 000 | 1,95 |
| 40 | 9,75 | 300 | 3,56 | 2 000 | 1,38 |
| 50 | 8,72 | 400 | 3,08 | 5 000 | 0,87 |
| 60 | 7,96 | 500 | 2,76 | 10 000 | 0,62 |
| 70 | 7,37 | 600 | 2,52 | 50 000 | 0,28 |
| 80 | 6,90 | 700 | 2,33 | 100 000 | 0,20 |
| 90 | 6,50 | 800 | 2,18 | 1 000 000 | 0,06 |
| 100 | 6,17 | 900 | 2,06 | 10 000 000 | 0,02 |

Mivel az elemi statisztikatankönyvekben a nagymintáról szóló megjegyzések a mintaátlag becslésénél szerepelnek, ezért megvizsgálhatjuk az ottani módszertant. Ha egy elméleti nagymintából indulunk ki, amelyre a normális eloszlás és a centrális határeloszlás tétele igaznak feltételezhető, akkor a populáció átlaga az

$$\hat{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}} \quad (12)$$

intervallumban lesz $(1 - \alpha)$ valószínűséggel, ahol \hat{x} jelenti a pontbecslést, és s annak szórását.

A következő lépés, hogy amennyiben kicsi a mintánk, de tudjuk, hogy az alapeloszlás normális, ismeretlen szórással, akkor az empirikus szórással standardizált mintaátlag Student-féle t -eloszlást fog követni (*Wackerly és szerzőtársai* [2008] 426. o.). Így a (12) egyenletben a z -értékeket a t -eloszlásból kapott t kritikus értékekre cseréljük. A t -értékek a szabadságfok növekedésével közelítenek a z -értékekhez. *Heumann–Schomaker* [2017] kiemeli, hogy az általa adott hüvelykujjszabály azon a felismerésen alapszik, hogy a t -eloszlás harminc szabadsági fok mellett nagyon közel kerül a standard normális eloszláshoz. Egy ezerelemű mintára a két érték már valóban egy 0,1 százalék széles intervallumon belül helyezkedik el. Viszont egy harmincelemű

mintára még közel 4 százalékos eltérés adódik közöttük. Összességében tehát azt tapasztaljuk, hogy azt a feltevést, amely szerint harmincelemű mintára a centrális határeloszlás tétele már jól alkalmazható, nem tudjuk megindokolni.

Statisztikai körökben egy historikus okokon alapuló magyarázat is elterjedt. Mivel a fentebb leírt módszertan jóval megelőzi a számítógépek elterjedését, így a becslésekhez sokáig papírra nyomtatott z - és t -táblákat használtak. A kényelmes használhatóság kedvéért célszerű volt csak egy-egy oldalt fenntartani a különböző tábláknak. Viszont egy oldalra megközelítően harminc sor fért ki, így egy sokkal inkább praktikus, mintsem elméleti ok miatt terjedhetett el ez a küszöbérték.

A brexitet megelőző közvélemény-kutatások esetében azonban egy-két ezres mintákról beszélhetünk. Láthatjuk, hogy kétezer fős minta esetén is a mintaátlag eloszlásfüggvénye és a standard normális eloszlásfüggvény között jelentős különbség adódhat.⁹ Így tehát a (9) képletben elkövethetünk egy hibát, ha a normális eloszlásból származó kritikus értéket használjuk. Megoldás lehetne, ha valamilyen más kritikus értéket használnánk, de annak meghatározásához a Berry–Esseen-tétel nem elég. A gyakorlatban ilyenkor *bootstrap* technikát alkalmaznak.

A (9) egyenlet levezetésekor azonban elkövettünk egy másik hibát is. Ha a szórás meghatározó (3) egyenletben elhagyjuk az alacsony populáció miatti korrigáló tényezőt, akkor a (4) egyenlet a következő alakú lesz:

$$\mathbb{P}\left[-z_{\alpha/2} < \frac{p - P}{\sqrt{P(1-P)/n}} < +z_{\alpha/2}\right] = 1 - \alpha, \quad (13)$$

amiből a következő intervallumbecslést kapjuk:

$$p_0 = p \pm z_{\alpha/2} \sqrt{P(1-P)/n}. \quad (14)$$

A Wald-módszernél ennél a pontnál feltettük, hogy $P = p$. Az így kapott konfidencia-intervallum azonban nem lesz pontos: a valódi lefedési szint nem fogja elérni a névleges szintet.

A konfidenciaintervallum becslésénél az általános cél a következőképpen fogalmazható meg: $\mathbb{P}(p \in CI) = 1 - \alpha$, ahol $\mathbb{P}(p \in CI)$ lefedési valószínűség és CI jelenti a konfidenciaintervallumot. Ha a mintaelemszám tart a végtelenbe, akkor aszimptotikusan a fenti egyenlőség teljesülni fog, azonban számos (n, p) párosításra ez nem fog fennállni (*Brown és szerzőtársai* [2001]). Így például a gyakran megadott $np > 10$ és $n(1-p) > 10$ kritériumok sem segítenek, hiszen azok inkább a centrális határeloszlás tételének a teljesülését próbálják biztosítani, így a Wald-féle közelítés hibáját nem korrigálják. Tehát hiába próbáltunk 95 százalékos intervallumot becsülni, elképzelhető, hogy csak 80 százalékosat kaptunk. Érdekes, hogy számos cikk erősen bírálja a Wald-módszer bevezető statisztikakönyvekben történő bemutatását. Véleményük szerint a módszer a fentebb említett komoly hiányossága miatt alapvető statisztikai eszközként nem ideális.¹⁰ Mivel ez a hiba

⁹ A 3. táblázat alapján egyenletes eloszlás esetén a különbség 1,38 százalék.

¹⁰ A módszer hiányosságaival *Agresti–Coull* [1998], *Brown és szerzőtársai* [2001] és *Newcombe* [2012] foglalkozik.

egyszerűen a $P=p$ feltetelezésből következik, ezért most vessük el azt. A (13) egyenletben az intervallum széleit felhasználva megadhatunk egy P -ben másodfokú egyenletet:

$$P^2 - 2pP + p^2 = z_{\alpha/2}^2 (P - P^2) / n, \tag{15}$$

amelyet átrendezve az

$$(1 + z_{\alpha/2}^2 / n)P^2 - (2p - z_{\alpha/2}^2 / n)P + p^2 = 0 \tag{16}$$

egyenletet kapjuk.

A másodfokú egyenletet egyszerűen megoldva a (17) konfidenciaintervallumot kapjuk:

$$\begin{aligned} p_0 &= \frac{p + z_{\alpha/2}^2 / 2n}{1 + z_{\alpha/2}^2 / 2n} \pm z_{\alpha/2} \frac{\sqrt{p(1-p)/n + z_{\alpha/2}^2 / 4n^2}}{1 + z_{\alpha/2}^2 / n} = \\ &= \hat{p} \pm z_{\alpha/2} \frac{\sqrt{p(1-p)/n + z_{\alpha/2}^2 / 4n^2}}{1 + z_{\alpha/2}^2 / n}. \end{aligned} \tag{17}$$

Ez az úgynevezett Wilson-féle konfidenciaintervallum (*Wilson* [1927]). A konfidenciaintervallum a hagyományos pontbecslés körül nem szimmetrikus, hiszen az új pontbecslés egyfajta súlyozott átlagként kerül meghatározásra. Viszont vegyük észre, hogy ha n valóban nagy, akkor $\hat{p} \approx p$.

Ha itt akarjuk rögzíteni a konfidenciaintervallum hosszát egy adott w szinten, akkor az ahhoz szükséges mintanagyságot a következőképpen adhatjuk meg:

$$N = \frac{2z_{\alpha}^2 p(1-p) - z_{\alpha/2}^2 w^2 \pm \sqrt{4z_{\alpha/2}^4 p(1-p)[p(1-p) - w^2] + w^2 z_{\alpha/2}^4}}{w^2}. \tag{18}$$

Így például a brexit esetében 95 százalékos konfidenciaszinten 1 százalékos hibatoleranciával több mint 9000 fős mintára lett volna szükség.

A Wilson-módszer teljesítményét számos tanulmány vizsgálja,¹¹ és jellemzően előnyben részesítik a korábban ismertetett Wald-féle konfidenciaintervallummal szemben. Ne feledkezzünk meg azonban arról, hogy itt is kihasználtuk a normális eloszlás feltételezését. Hogy még pontosabb legyen a becslés, a haladó statisztikatan-könyvek Wald és Wilson módszerei helyett – *Clopper-Pearson* [1934] munkája után – a Clopper-Pearson-módszert javasolják. A módszer a következőképpen becsüli a konfidenciaintervallum végpontjait:

$$\sum_{k=x}^n \binom{n}{k} p_0^k (1-p_0)^{(n-k)} = \alpha/2, \tag{19}$$

$$\sum_{k=0}^x \binom{n}{k} p_0^k (1-p_0)^{(n-k)} = \alpha/2, \tag{20}$$

¹¹ A populációarány lehetséges becslésének számos módszerét vizsgálja *Krishnamoorthy-Peng* [2007], *Agresti-Coull* [1998] vagy *Brown és szerzőtársai* [2001] is.

ahol x a mintában az adott kimenetek száma, és p_0 a konfidenciaintervallum alsó vagy felső határa. Azaz normális eloszlás feltételezése nélkül pontos konfidenciaintervallumot definiálunk. A (19) és (20) egyenlet alapján a lefedési szint definiálva van a nominális értéknél, így minden esetben elérjük azt, sőt sokszor felül is becsüljük. Ha $x \notin \{0, n\}$, a konfidenciaintervallum a következőképpen számolható:

$$\left[1 + \frac{n-x+1}{xF_{2x, 2(n-x+1), 1-\alpha/2}} \right]^{-1} < p < \left[1 + \frac{n-x}{(x+1)F_{2(x+1), 2(n-x), \alpha/2}} \right]^{-1}, \tag{21}$$

ahol $F_{x_1, x_2, y}$ az F -eloszlás x_1, x_2 szabadságfokkal és y szignifikanciaszinttel. Az így kapott konfidenciaintervallumot számos kritika érte, mivel az a pontos becslés miatt szélesebb a más módszerek alapján kapottaknál. Ennek megfelelően jellemzően nagyobb mintára is van szükség, mint az előző módszerek esetében. *Thulin* [2014] munkája közelítést ad a Clopper–Pearson-módszer alapján számolt konfidenciaintervallumhoz szükséges mintaszámra.

$$n = \frac{2z^2 p(1-p) + 2z\sqrt{z^2 p^2(1-p)^2 + wp(1-p)} + w}{w^2}, \tag{22}$$

ahol p egy *a priori* sejtés a P értékére, és w az intervallum várható hossza. Az így kapott minta valóban nagyobb lesz, de az eltérés nem jelentős.

A 4. és az 5. táblázatban mutatjuk a minta szükséges nagyságát különböző konfidenciaszintek és hibanaagságok mellett, miközben a $p(1-p)$ szorzat helyére annak maximumát (0,25) helyettesítettük. Láthatjuk, hogy a Clopper–Pearson-módszer jellemzően magasabb mintaszámot vár el. A különbség aránya azonban a pontosság növekedésével párhuzamosan csökken. Így például 10 százalék széles konfidenciaintervallum mellett

4. táblázat

A szükséges mintanagyság a Wilson-módszerrel számolva

| Konfidenciaintervallum hossza (százalék) | Konfidenciaszint | | |
|---|------------------|--------|--------|
| | 90 | 95 | 99 |
| | százalék | | |
| 10 | 265 | 376 | 650 |
| 9 | 329 | 467 | 806 |
| 8 | 417 | 593 | 1 023 |
| 7 | 547 | 776 | 1 341 |
| 6 | 746 | 1 059 | 1 830 |
| 5 | 1 077 | 1 529 | 2 641 |
| 4 | 1 686 | 2 393 | 4 134 |
| 3 | 3 001 | 4 261 | 7 359 |
| 2 | 6 758 | 9 596 | 16 574 |
| 1 | 27 050 | 38 407 | 66 336 |

5. táblázat

A szükséges mintanagyság a Clopper–Pearson-módszerrel számolva

| Konfidenciaintervallum hossza (százalék) | Konfidenciaszint | | |
|---|------------------|--------|--------|
| | 90 | 95 | 99 |
| | százalék | | |
| 10 | 290 | 404 | 683 |
| 9 | 356 | 496 | 841 |
| 8 | 447 | 625 | 1 062 |
| 7 | 580 | 812 | 1 382 |
| 6 | 785 | 1 100 | 1 876 |
| 5 | 1 122 | 1 576 | 2 694 |
| 4 | 1 741 | 2 451 | 4 197 |
| 3 | 3 072 | 4 335 | 7 439 |
| 2 | 6 863 | 9 703 | 16 687 |
| 1 | 27 255 | 38 614 | 66 549 |

még több mint 5 százalékos eltérés adódik a mintaszámokban, addig ez 1 százalék szélesség mellett már kevesebb mint 1 százalékos eltérést eredményez.

Ahogy azt már korábban említettük, a brexit előtt elvégzett közvélemény-kutatások egy-két ezres mintát használtak. Ugyan elsőre hihetetlennek tűnik, hogy egy több mint 65 milliós populációról kétezer megfigyelés alapján állításokat fogalmazzunk meg, a 4. és 5. táblázat alapján mégis azt látjuk, hogy ilyen kis elemszámú mintával is relatíve alacsony hibát érhetünk el. Azonban fontos szem előtt tartani, hogy ez a hiba még nagy ahhoz, hogy egy szoros választás kimenetéről dönteni lehessen, hiszen a pontos eredményhez az eredeti minta sokszorosára van szükség. Ennél a pontnál szeretnénk kiemelni, hogy ez nem jelenti azt, hogy a pontos becsléshez a lakosság jelentős részének megkeresésére lett volna szükség. Még az általunk javasolt mintanagysággal számolva is a teljes lakosság 0,05 százalékáról beszélhetünk. Egy ilyen jellegű felmérés sem a technikai megvalósítás, sem pedig a költségek szempontjából nem tekinthető elérhetetlennek. A következőkben a statisztika eszköztárának egy alternatív elemét választjuk az elemzéshez.

Hipotézisvizsgálat

Az eddigi elemzést a konfidenciaintervallum becslésének módszertanára építettük. A cél az volt, hogy a minta alapján előállítsunk egy szűk intervallumot, amely megfelelő valószínűség mellett lefedi a tényleges populációs arányt. Valójában egy közvélemény-kutatás során felmerült kérdést egyszerűbben, hipotézisvizsgálattal is megválaszolhatunk.¹² A hipotézisvizsgálat és a konfidenciaintervallum becslése

¹² Köszönet a bírálónak megjegyzéséért.

ugyan szorosan összekapcsolódó témák, mégis ebben a részben a hipotézisvizsgálaton alapuló megközelítést választjuk.

Szemben a konfidenciaintervallum becslésével, tegyük fel azt az egyszerű kérdés, hogy Trump támogatottsága nagyobb-e, mint Hillary Clintoné, vagy éppen a kilépéspártiak vezetnek-e. A hipotézisvizsgálat megfelelő kivitelezését például a biostatistikai szakirodalma dolgozza fel.¹³ A fenti kérdéseket a szakirodalom alapján a felsőbbrendűségi hipotézisvizsgálat keretébe lehet beilleszteni. Az egy mintára alapuló felsőbbrendűségi hipotézisvizsgálatot *Chow és szerzőtársai* [2007] alapján a következőképpen formalizálhatjuk.

Legyen P a valódi mintaarány és p_0 egy referenciaérték. Ha például P jelenti a kilépéspártiak arányát, akkor a

$$H_0: P = p_0, H_1: P > p_0 \quad (23)$$

hipotézisekkel tesztelhetjük, hogy a kilépéspártiak vezetnek-e, feltéve hogy a $p_0 = 0,5$ kikötéssel élünk. Hipotézisvizsgálat során két hibát véthetünk. Az elsőfajú hiba esetén elutasítjuk a H_0 hipotézist, miközben az igaz volt. Jelöljük ennek a hibának a valószínűségét a szakirodalmat követve α -val. A második hibát akkor követjük el, ha nem utasítjuk el a H_0 hipotézist, amikor a H_1 az igaz. Ha az ilyen típusú, másodfajú hibához tartozó valószínűséget β -val jelöljük, akkor a teszt erejét $(1 - \beta)$ adja meg. A teszt ereje tehát annak a valószínűsége, hogy elutasítjuk a H_0 hipotézist, feltéve, hogy H_1 az igaz. Mivel a H_1 alternatív hipotézishez számos P -érték is tartozhat, ezért a β sem lesz egyértelmű.

Chow és szerzőtársai [2007] vagy *Devore* [2015] munkáit követve a fenti H_0 hipotézis melletti tesztstatisztikát a következőképpen írhatjuk fel:

$$Z = \frac{\sqrt{n}(p - p_0)}{\sqrt{p_0(1 - p_0)}}. \quad (24)$$

Nagyminta esetén Z megközelítően standard normális eloszlást követ. Így a H_0 hipotézist α szignifikanciaszint mellett elutasítjuk, ha

$$Z > z_\alpha. \quad (25)$$

A teszt erejének meghatározásához szükségünk van a β -értékre. Mivel a H_1 hipotézis mellett a Z tesztstatisztika már nem standard normális, ezért azt újra kell standardizálnunk. A H_1 alatt azonban P számos értéket felvehet, amelyek mellett a β -t specifikusan kell meghatározni. Ezért most specifikáljuk P -t, és éljünk a $P = p' > p_0$ feltételezéssel. Ebben az esetben a teszt ereje felírható a (26) alakban:

¹³ A témában népszerű tankönyveknek számítanak *Rosner* [2015] vagy *Chow és szerzőtársai* [2007].

$$\begin{aligned}
 1 - \beta &= \mathbb{P}[H_0 - t \text{ elutasítom} | H_1 \text{ az igaz}] = \mathbb{P}\left[p > p_0 + z_\alpha \sqrt{p_0(1-p_0)/n} \mid H_1 \text{ az igaz}\right] = \\
 &= \mathbb{P}\left[\frac{p - p'}{\sqrt{p'(1-p')/n}} > \frac{p_0 - p' + z_\alpha \sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}} \mid H_1 \text{ az igaz}\right] = \\
 &= 1 - \Phi\left[\frac{p_0 - p' + z_\alpha \sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}\right]. \tag{26}
 \end{aligned}$$

A (26) egyenlet megadja, hogy bizonyos paraméterezés mellett mekkora lesz a teszt ereje, valamint a másodfajú hiba valószínűsége. Ezt azonban meg is fordíthatjuk, és rögzített β mellett megadhatunk egy egyenletet a minta elemszámára. Hiszen a (26) egyenlet alapján

$$\beta = \Phi\left[\frac{p_0 - p' + z_\alpha \sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}\right], \tag{27}$$

amelyből

$$-z_\beta = \frac{p_0 - p' + z_\alpha \sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}. \tag{28}$$

A (28) egyenletből n értékét kifejezve, megkapjuk a mintaelemszám – a teszt erejének elemzéseként kapott – formuláját:

6. táblázat

Mintanagyság az első- és a másodfajú hibák specifikálása mellett

| β | α | | |
|----------|----------|--------|--------|
| | 0,01 | 0,05 | 0,1 |
| $P=0,53$ | | | |
| 0,05 | 4 371 | 3 001 | 2 375 |
| 0,1 | 3 607 | 2 374 | 1 822 |
| 0,2 | 2 780 | 1 713 | 1 249 |
| $P=0,52$ | | | |
| 0,05 | 9 847 | 6 758 | 5 349 |
| 0,1 | 8 127 | 5 348 | 4 103 |
| 0,2 | 6 265 | 3 860 | 2 815 |
| $P=0,51$ | | | |
| 0,05 | 39 417 | 27 050 | 21 406 |
| 0,1 | 32 534 | 21 405 | 16 420 |
| 0,2 | 25 083 | 15 452 | 11 267 |

$$n = \left[\frac{z_{\beta} \sqrt{p'(1-p')/n} + z_{\alpha} \sqrt{p_0(1-p_0)/n}}{p' - p_0} \right]^2. \quad (29)$$

A 6. táblázatban ismertetjük a (29) egyenlettel kapott mintaelemszámokat különböző α és β feltételek mellett. Láthatjuk, hogy egy apró előnyt (51 százalék) kimutató hipotézisvizsgálathoz több tízezres mintára van szükségünk, ha mind a teszt erejét, mind a szignifikanciaszintjét előre rögzítjük. Magasabb arány (53 százalék) teszteléséhez elegendő lehet a néhány ezer fős minta, de szoros verseny esetén az így kapott eredmények nem fognak a H_0 hipotézis elvetéséhez vezetni. Így a közvélemény-kutatás igazán nem tud lényegi megállapítással élni.

Következtetések

A fentiek tükrében megállapíthatjuk, hogy a brexitet egy óriási félreértés előzte meg. Néhány kivételtől eltekintve a közvélemény-kutatások brit bennmaradást hoztak ki eredményül. A média a pontbecslésre támaszkodva felerősítette ezt a tévképzetet, és erre a kimentre készítette fel a lakosságot és a piacokat. A tényleges eredmény azonban szinte sohasem az Európai Unióban maradás volt, hanem az, hogy nem tudják, mi lesz a kimenet. De miért is nem ezt kommunikálták?

A legegyszerűbb ok feltételezhetően az ilyen típusú intézmények megbízási jellegű tevékenysége. Egy jelentős költségeket finanszírozó megbízó választ vár a közvélemény-kutató intézettől. Így az az eredmény, amely szerint nincs válasz, nem elfogadható. Ahogy fentebb láttuk, a kis minta nagy hibát jelent, aminek a tükrében különösen meglepő, hogy gyakorlatilag minden közvélemény-kutatás az EU-ban maradást hozta ki eredményül. Nagy hiba mellett ugyanis azt várnánk, hogy az eredmények szóródni fognak az egyik és a másik oldal körül. Ez a megfigyelés azonban megkérdőjelezi azt, hogy a közvélemény-kutatók ténylegesen becsülték-e az eredményeket, vagy inkább az adatok begyűjtése után valamilyen szakértői kiigazítást alkalmaztak. Ennek egyik eszköze lehet a súlyozás, ami kellően nagy szabadságfokot ad a közvélemény-kutatóknak.

Ilyen jellegű eredmény becslésekor a felmérést végzők nem csak a mintára támaszkodnak. A kérdésekre választ adó személyeket gyakran szociális és társadalmi jellemzők alapján csoportosítják, azzal a feltételezéssel élve, hogy a csoportok homogénebben viselkednek, mint a teljes minta. A csoportokat nézve azonban rendelkezésre állnak népszámlálásokból származó adatok, amelyek segítenek meghatározni, hogy az egyes csoportoknak mennyire is lesz szignifikáns szerepük a tényleges választáskor. Így tehát a közvélemény-kutatók kiegészítik az általuk elérhető információit, annak érdekében, hogy pontosabb becslést adjanak. Fontos hangsúlyozni, hogy ilyenkor a teljes minta több, kisebb elemszámú almintára esik szét. A kisebb elemszám viszont, ahogy azt az előzőekben megmutattuk, nagyobb hibát eredményez. Tehát ha eredetileg is 3–5 százalék hibával dolgoztak, akkor az alcsoportokra számított hiba az eredeti érték többszörösét elérheti. Ebből a szempontból az ORB

International adattáblái két ok miatt tanulságosak. Egyrészt a csoportokhoz tartozó hiba feltüntetését csak a brexit után tették meg, másrészt ezek nagysága akár a 10 százalékos értéket is eléri.

Ez részben elvezet a második okhoz, amely a nyilvánosságra hozatalt érinti. A közvélemény-kutatók honlapjain néhány esetben elérhető egy módszertani leírás. A legtöbbször ez az adatgyűjtési folyamatra koncentrál, azaz arra, hogy telefonon vagy személyesen keresték-e fel a válaszadókat, valamint, hogy kiket is vontak be a mintába. A módszertan sajnos egyik esetben sem részletezi, hogy mekkora hibával dolgoznak, és azt milyen alapon határozzák meg, vagy milyen súlyozási módszert alkalmaztak a nyers adatokra. Ennek pedig az az oka, hogy az ilyen intézmények módszertana nincsen szabályozva. A brit közvélemény-kutató intézmények egyfajta védjegyként gyakran a British Polling Council nevét húzzák elő. Ez valójában a közvélemény-kutató intézmények által létrehozott csoport, amely minimális kritériumokat állított fel a közvélemény-kutatások eredményeinek közzétételére.¹⁴ A British Polling Council által közzétett dokumentumban szereplő követelmények között nem szerepel a mintavételi hiba megadása.

A British Polling Council honlapján azonban közzétesz egy „gyakran feltett kérdések” jellegű leírást, amelynek 11. pontjában a következő olvasható: „A közvélemény-kutatások nem tökéletesek, de ezek kínálják a legjobb vagy a legkevésbé rossz utat a nyilvánosság véleményének felméréséhez. [...] Ha egy jól tervezett, reprezentatív felmérés azt találja, hogy a közvélemény 70 : 30 százalék arányban oszlik meg egy kérdésen, akkor még egy 10 százalékos hiba sem változtatja meg, hogy az egyik nézet sokkal többször fordul elő, mint a másik. Egy szoros választási küzdelem esetén azonban az egyik fél 5 százaléknál kisebb előnye (egy-két ezer fős minta alapján) nem tekinthető biztos jelzésnek arra nézve, hogy ki is vezet a felmérés készítésekor, nem hogy ki fog napok, hetek és hónapok múlva.”¹⁵ Az alacsony mintaszám önmagában még nem feltétlenül okoz teljes félreértést. Amikor azonban az eredmény megfelelő értelmezéséhez szükséges, a fentihez hasonló leírások, mondhatni, az „apró betűs részben” vannak elrejtve, és egy-egy új eredmény szinte a másikkal versengve próbál minél szélesebb körhöz eljutni, anélkül hogy azt tisztán, egyértelműen közölnék, akkor hatalmas problémákkal nézhetünk szembe. Úgy gondoljuk, hogy ha a közzétett adattáblák első oldalán egyszerűen feltüntették volna a becsült konfidenciaintervallum alsó és felső határát, a brexit által okozott meglepetés sokkal kisebb lett volna. Ezt a közvélemény-kutatóknak érdemes megfontolni, mert a legutóbbi kudarcaik után nagyon mélyről kell újraépíteniük a hitelességüket.

Érdemes több mint egy év távlatából áttekinteni az azóta történt eseményeket. A brexit és az amerikai elnökválasztás után az internetet elárasztották a közvélemény-kutatók munkáját elemző bejegyzések. A hibákból levont következtetések a statisztikai mintavétel nehézségeinek szinte minden lehetséges problémáját lefedték. A közhangulat azonban egységesen a közvélemény-kutatók bukásaként ítélte meg az eseményeket.

¹⁴ <http://www.britishpollingcouncil.org/statement-of-disclosure>.

¹⁵ <http://www.britishpollingcouncil.org/a-journalists-guide-to-opinion-polls/#q2>.

A problémát az AAPOR¹⁶ is felismerte, és azonnali vizsgálatot indított az elnökválasztás során lefolytatott közvélemény-kutatások hibáinak feltérképezésére. A végső riport (AAPOR [2017]) alapján a Trump-támogatottság alulbecslésének három leginkább bizonyítható oka a következő: a preferenciák változása a kampány utolsó hetében, a „szégyenlős” Trump-szavazók magas száma, valamint az alacsony és magas iskolai végzettséggel rendelkezők rossz súlyozása a mintákban.

A fenti elemzés jellemzően a közvélemény-kutatóktól független hibákat nevez meg. Ugyan sok fejtegetés egyetért abban, hogy a megkérdezettek között a Trump-preferenciát fel nem vállalók száma magas volt, mégis a nem közvélemény-kutatók által készített elemzésekben más szempontok is feltűnnek. Többek között az általunk javasolt mintaelemszám-problémát is többen előtérbe helyezik.¹⁷ A közvélemény-kutatóktól független hibákat azonban már az újabb tapasztalatok alapján is nehéz elhinni. 2017. április 18-án Theresa May brit miniszterelnök 2017. június 8-ra előrehozott választásokat rendelt el, amivel nem titkolt célja pozíciójának megerősítése volt a brexittárgyalások véghezvitele érdekében. A döntés helyességét kevesen kérdőjelezték meg, hiszen az akkori közvélemény-kutatások 20 százalékpontos különbséget mutattak a Konzervatív Párt és a Munkáspárt között. Ahogy azt ma már tudjuk, a tényleges kimenet szerint a Konzervatív Párt elvesztette a többségét, és koalícióra kényszerült. A 2017-es választás előtt sokan felelevenítették a 2015-ös választásokat, ahol a közvélemény-kutatók nagyobb része a vesztes felet hozta ki győztesül.¹⁸

A fentiek tükrében feltehetjük magunknak a kérdést, hogy ha a közvélemény-kutatók mostantól csakis 10 000 fő feletti mintával dolgoznának, akkor vajon tökéletesek lennének-e az előrejelzések? A válasz egyszerűen nem. A közvélemény-kutatás egy nagyon komplex feladat, amelyet rengeteg torzító tényező nehezít. Az ilyen intézmények nem csak a minta nagyságával tudnak hibázni. Jól illusztrálja ezt *Hahn-Meeker* [1991] gondolata: „... az adatmennyiség problémája gyakran másodlagos az adatminőséggel szemben. [...] Másképpen mondva, a mintaelemszám emelése önmagában javítja a becslés precizitását, de nem szükségképpen annak pontosságát.” (141. o.) *Weisberg* [2009] a teljeshiba-alapú megközelítés (*total survey error approach*) egyik átfogó munkája. *Weisberg* szerint a mintavételezési hiba csak a „jéghegy csúcsa”. Mivel a hiba felére csökkentéséhez a minta nagyságát közel négyszeresére kell növelni, ezért ez nagy költségekkel jár. A teljeshiba-alapú megközelítés szerint ezeket a költségeket érdemesebb más hiba csökkentésére fordítani.

Jelen munka tárgya ugyanúgy a kérdésfelvetés, mint a válaszadás. Nem állítjuk, hogy a brexitet vagy a Trump győzelmét megelőző közvélemény-kutatások sikertelenségének egyetlen oka a mintanagyság téves megválasztása volt. Így tehát azt sem,

¹⁶ Közvélemény-kutatások Amerikai Egyesülete (American Association for Public Opinion Research) – a közvélemény-kutatások fejlődéséért létrehozott intézmény.

¹⁷ A nagyobb minták mellett érvelő nem tudományos cikkekre példák: *Wilson* [2016] vagy *Roberts* [2016].

¹⁸ Az akkori eredmények elérhetők: <http://www.bbc.com/news/uk-politics-32751993>, míg a 2017-es választás előtt a 2015-ös közvélemény-kutatások hibáinak áttekintése elérhető: <https://www.significance magazine.com/politics/549-what-s-happened-to-the-polls-since-the-2015-uk-election#sec13>.

hogy a korábban felsorolt faktorok, mint például a rejtett, illetve változó preferenciák, a reprezentativitás hiánya vagy a bizonytalanság nem befolyásolták a közvélemény-kutatók eredményeit. Mi több, elismerjük, hogy ezek hatása legalább annyira fontos lehet, mint a mintanagyság hibás választása. Tanulmányunk célja azonban mégis az volt, hogy megvitassuk a kevés kvantitatív úton specifikálható paraméter közül az egyik – a mintaelemszám – kérdését. A mintanagyság ugyanis egy olyan alapvető mennyiség, amelynek a megfelelő szintre történő beállítása elengedhetetlen. Ez a megállapítás azonban nemcsak a közvélemény-kutatásokra vonatkozik, hanem számos más becslésre is, és ennek hátterében a centrális határeloszlás tétele húzódik meg. Így talán nem túlzás azt állítani, hogy az általunk vizsgált kérdéskör a statisztika, az ökonometria és az idősorelemzés eszköztárának és modelljeinek jelentős része szempontjából fontos lehet. Ezért válaszadás előtt mindig tegyük fel a kérdést, hogy mekkora is a minta?

Hivatkozások

- AAPOR [2017]: An Evaluation of 2016 Election Polls in the U.S. Ad Hoc Committee on 2016 Election Polling. American Association for Public Opinion Research, <http://www.aapor.org/getattachment/Education-Resources/Reports/AAPOR-2016-Election-Polling-Report.pdf.aspx>.
- AGRESTI, A.–COULL, B. A. [1998]: Approximate is better than “exact” for interval estimation of binomial proportions. *The American Statistician*, Vol. 52. No. 2. 119–126. o. <https://doi.org/10.1080/00031305.1998.10480550>.
- ASHER, H. [2011]: *Polling and the public: What every citizen should know*. SAGE, London.
- BARTLETT, J. E.–KOTRLIK, J. W.–HIGGINS, C. C. [2001]: Organizational research: Determining appropriate sample size in survey research. *Information Technology, Learning, and Performance Journal*, Vol. 19. No. 1. 43–50. o. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.486.8295&rep=rep1&type=pdf>.
- BROWN, L. D.–CAI, T. T.–DASGUPTA, A. [2001]: Interval estimation for a binomial proportion. *Statistical Science*, Vol. 16. No. 2. 101–117. o. <https://doi.org/10.1214/ss/1009213286>.
- CHOW, S.-C.–WANG, H.–SHAO, J. [2007]: *Sample size calculations in clinical research*. Chapman and Hall/CRC Biostatistics Series, CRC Press, <https://doi.org/10.1201/9781584889830>.
- CLOPPER, C. J.–PEARSON, E. S. [1934]: The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, Vol. 26. No. 4. 404–413. o. <http://dx.doi.org/10.1093/biomet/26.4.404>.
- COCHRAN, W. G. [1977]: *Sampling techniques*. John Wiley and Sons, New York.
- DEVORE, J. L. [2015]: *Probability and Statistics for Engineering and the Sciences*. Cengage Learning, <https://doi.org/10.2307/2532427>.
- D. Sz. [2016]: Orbán Viktor is bezáll a brit EU-népszavazási kampányba. Index, június 19. http://index.hu/belfold/2016/06/19/orban_viktor_is_beszall_a_brit_eu-nepszavazasi_kampanyba.
- EDSALL, T. B. [2016]: How Many People Support Trump but Don’t Want to Admit It? *The New York Times*, május 11. <https://www.nytimes.com/2016/05/11/opinion/campaign-stops/how-many-people-support-trump-but-dont-want-to-admit-it.html>.
- FRY, J. [2016]: A statistical reaction to Brexit Features. *StatisticsViews*, augusztus 11. <http://www.statisticsviews.com/details/feature/9690531/A-statistical-reaction-to-brexit.html>.

- HAHN, G. J.–MEEKER, W. Q. [1991]: *Statistical intervals: A guide for practitioners*. Wiley, New York, <https://doi.org/10.1002/9780470316771>.
- HEUMANN, C.–SCHOMAKER, M. [2017]: *Introduction to Statistics and Data Analysis: With Exercises, Solutions and Applications in R*. Springer, <http://dx.doi.org/10.1007/978-3-319-46162-5>.
- KREJCIE, R. V.–MORGAN, D. W. [1970]: Determining sample size for research activities. *Educational and Psychological Measurement*, Vol. 30. No. 3. <http://dx.doi.org/10.1177/001316447003000308>.
- KRISHNAMOORTHY, K.–PENG, J. [2007]: Some properties of the exact and score methods for binomial proportion and sample size calculation. *Communications in Statistics—Simulation and Computation*, Vol. 36. No. 6. 1171–1186. o. <http://dx.doi.org/10.1080/03610910701569218>.
- LAPLACE, P. S. [1812]: *Théorie analytique des probabilités*. Courcier, Párizs.
- MEEKER, W. Q.–HAHN, G. J.–ESCOBAR, L. A. [1991]: *Statistical Intervals: A Guide for Practitioners and Researchers*. Wiley Online Library, <http://dx.doi.org/10.1002/9781118594841>.
- NEWCOMBE, R. G. [2012]: *Confidence intervals for proportions and related measures of effect size*. CRC Press, <http://dx.doi.org/10.1201/b12670>.
- REA, L. M.–PARKER, R. A. [2014]: *Designing and conducting survey research: A comprehensive guide*. John Wiley & Sons.
- ROBERTS, D. [2016]: The ‚failure’ of election polling was about 3 key things. Yahoo Finance, november 10. <https://finance.yahoo.com/news/the-failure-of-election-polling-was-about-3-key-things-182053428.html>.
- ROSNER, B. [2015]: *Fundamentals of biostatistics*. Nelson Education.
- SHEVTSOVA, I. G. [2014]: On the absolute constants in the Berry-Esseen-type inequalities. *Doklady Mathematics*, Vol. 89. No. 3. 378–381. o. <http://dx.doi.org/10.1134/S1064562414030338>.
- THULIN, M. [2014]: The cost of using exact confidence intervals for a binomial proportion. *Electronic Journal of Statistics*, Vol. 8. No. 1. 817–840. o. <http://dx.doi.org/10.1214/14-EJS909>.
- WACKERLY, D. D.–MENDENHALL, W.–SCHEAFFER, L. R. [2008]: *Mathematical Statistics with Applications*. Cengage Learning, <https://pdfs.semanticscholar.org/911e/69b44eac326eb20f790241782efa8edb499a.pdf>.
- WEISBERG, H. F. [2009]: *The total survey error approach: A guide to the new science of survey research*. University of Chicago Press, <http://dx.doi.org/10.7208/chicago/9780226891293.001.0001>.
- WEISS, N. [2017]: *Introductory Statistics*. Pearson, 10th Edition.
- WILSON, E. B. [1927]: Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, Vol. 22. No. 158. 209–212.o. <http://dx.doi.org/10.1080/01621459.1927.10502953>.
- WILSON, J. [2016]: Election Polling: A Failure of Small Data. <https://www.linkedin.com/pulse/election-polling-failure-small-data-joel-wilson>.
- ZUBOR ZALÁN [2016]: Komoly brexit-hátrányt jósol a legújabb felmérés. Origo, június 23. <http://www.origo.hu/gazdasag/20160623-komoly-brexit-hatranyt-josol-a-legujabb-felmeres.html>.