

COVID_19

2020. április 8. szerda, 10:40

Magyarországon 2020. április 6-án 11 fővel nőtt a fertőzöttek hivatalos száma

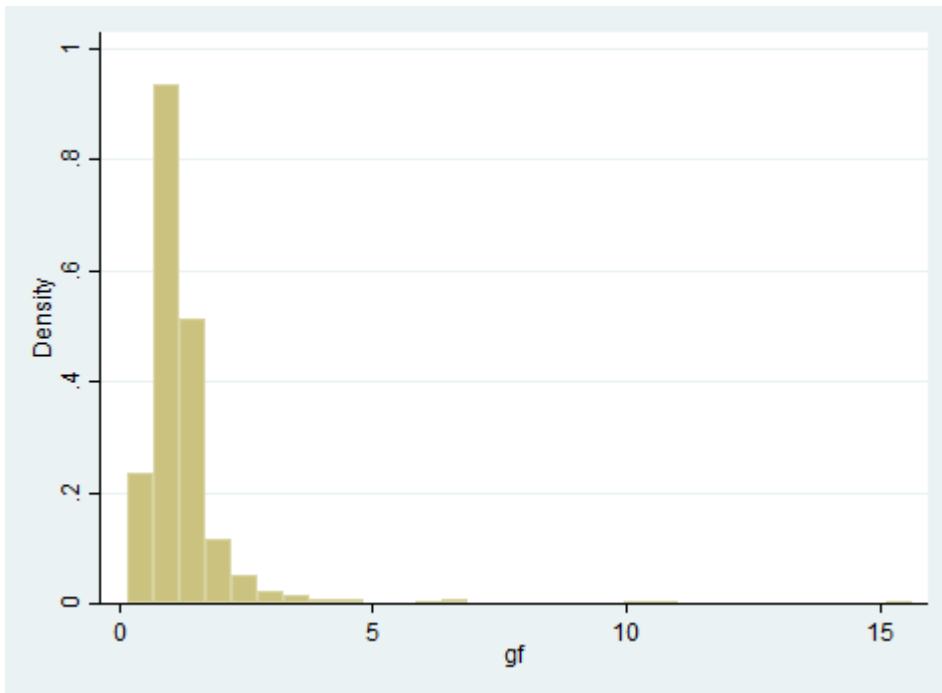
Magyarországon 2020. április 6-án 11 fővel nőtt a fertőzöttek hivatalos száma

2020. április 6-án 11 fővel nőtt a fertőzöttek hivatalos száma Magyarországon. Mivel az előző nap 55 fővel nőtt ez az szám, ezért a járvány terjedésének sebességét jelző növekedési tényező (*growth factor*, $\Delta N_t / \Delta N_{t-1}$) értéke erre a napra számítva 0,2 lett (11/55). Ez kivételesen alacsony érték, sőt, ha végignézzük Ausztria, Ausztrália, Csehország, Egyesült Királyság, Franciaország, Hollandia, Lengyelország, Németország, Norvégia, Olaszország, Spanyolország, Svájc, Svédország és USA megfelelő adatait¹ úgy, hogy csak azokat a napokat vesszük figyelembe, amikor a fertőzöttek száma (N) túlhaladta a százat, akkor azt kell látnunk, hogy ez a szám a második legalacsonyabb. Összesen 478 megfigyelésünk van, ennyi országra és napra számoljuk ki a növekedési tényezőt a fenti feltételekkel. A legalacsonyabb érték 0,16, egy svájci érték, amikor a fertőzöttek hivatalos száma ott 2.325 fő volt és egy nap alatt mindössze 136 fővel nőtt ez a szám.

Mi magyarázhatja ezt az alacsony értéket? Arról van szó, hogy április elejére már sikerült megállítani a fertőzöttek számának növekedését Magyarországon? Vagy inkább arról, hogy a fertőzöttek számbavételének procedúrája olyan, hogy nem minden nap egyforma: valamikor több, valamikor pedig kevesebb fertőzöttet képes detektálni a rendszer? Az utóbbira utal, hogy másnap már 73-ra ugrott az új fertőzöttek hivatalos száma. Ezzel pedig az erre a napra számított növekedési tényező 6,64-re nőtt (73/11). De az is lehet, hogy valamilyen adathibával állunk szemben. Természetesen lehet valamiféle szezonális a fertőzöttek számának alakulásában, ami magyarázhatja ilyen értéket. De ekkor időről időre meg kellene jelennie ennek a periodikus különbségnek a megfigyelt országok statisztikáiban – és az adatok eddigi vizsgálata ezt a feltételezést nem támasztja alá. Aztán az is igaz lehet, hogy a növekedési faktor értékeinél mutatkozó kiugró értékek gyakorisága összefügg a fertőzöttek hivatalos számával. Minél inkább előrehalad a járvány, azaz minél nagyobb már a fertőzöttek hivatalos száma, annál ritkábban fordulhatnak elő ilyen lassulások és ugrások.

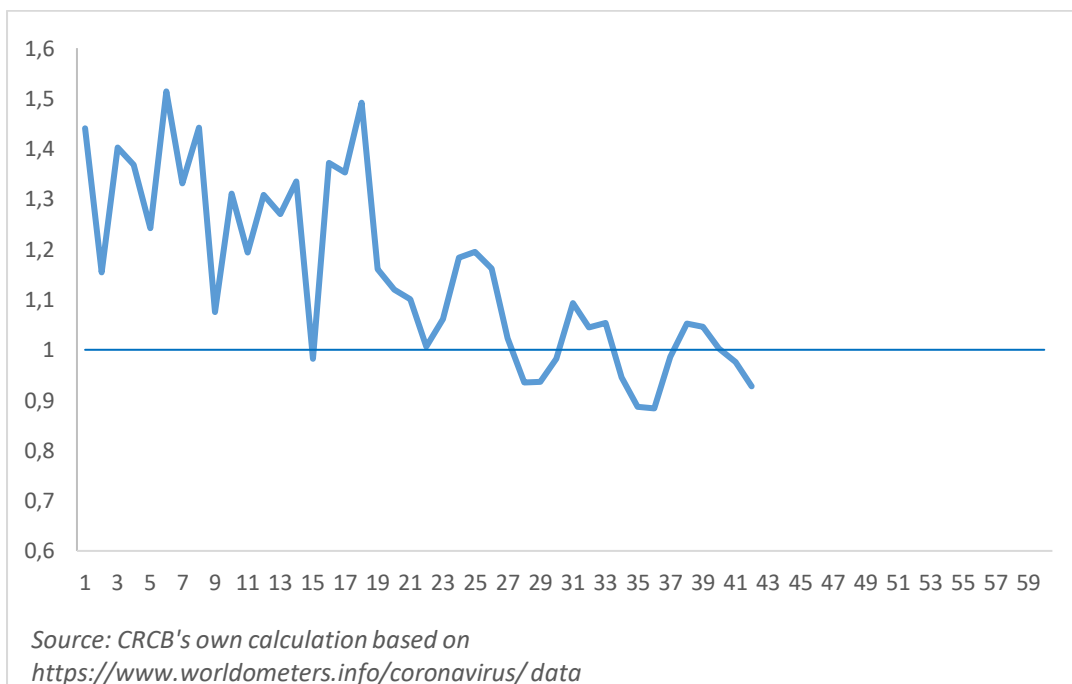
A növekedési tényező hisztogramját az 1. ábra mutatja N>100 esetében.

¹ A felhasznált adatok forrása: <https://www.worldometers.info/coronavirus/>



1. ábra: A növekedési tényező ($\Delta N_t / \Delta N_{t-1}$) hisztogramja, $N > 100$ esetében

A növekedési tényező időben csökkenő tendenciát mutat, ahogy az egyes országokban sikerül megfogni a járvány terjedését, értéke úgy közelít egyhez, és amint az új fertőzöttek hivatalos száma csökkenni kezd, értéke egy alá esik. Lásd erről a jól dokumentált olasz adatokat (2. ábra).



2. ábra: A növekedési tényező alakulása Olaszországban, három napos mozgó átlag attól a naptól fogva, hogy a fertőzöttek hivatalos száma meghaladta a 10 főt.

A növekedési tényező statisztikái a megfigyelt országokra és a megadott feltétel (N>100) mellett az alábbi (1. táblázat).

alsó 1%	0,3399
alsó 5%	0,6425
medián	1,0854
felső 5 %	2,4618
felső 1 %	6,6364
átlag	1,3073
szórás	1,1498
N	478

1. táblázat: A növekedési tényező statisztikái

Ha az alsó 5 százaléknál kisebb és a felső 5 százaléknál nagyobb értékeket tekintjük kiugrónak, akkor az ilyen esetek aránya csökken a fertőzöttek számának növekedésével (lásd a 2. táblázatot).

N	0	1	Total
< 500	77	13	90
< 1000	35	10	45
< 3000	78	11	89
< 5000	47	3	50
<10000	49	4	53
<50000	85	1	86
>=50000	59	6	65
Total	430	48	478

2. táblázat: Outlierek [1] előfordulása a fertőzöttek hivatalos száma (N) szerint

A növekedési tényezők kiugró értékeinek előfordulási gyakorisága országonként meglehetősen különböző (lásd a 3. táblázatot).

country	0	1	Total
AT	27	3	30
AUS	21	6	27
CH	24	9	33
CZ	21	4	25
DE	34	4	38
ES	35	2	37
FR	29	9	38
HU	16	2	18
IT	44	0	44
NL	32	0	32
NO	26	5	31
PL	23	1	24
SE	31	1	32
UK	32	1	33
USA	35	1	36
Total	430	48	478

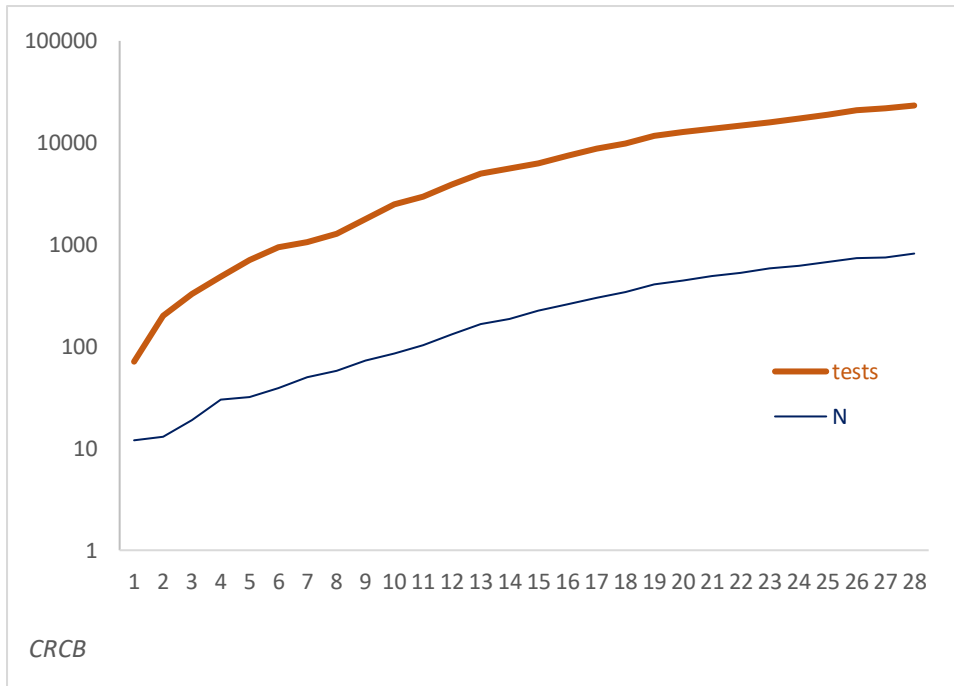
3. táblázat: Outlierek [1] előfordulása országonként

Érdekes, hogy a fertőzöttek hivatalos számának magas értéke mellett is előfordulnak kiugró értékek a növekedési tényezőnél, például Franciaországban. Itt, miután a fertőzöttek hivatalos száma 59.105 lett, egy nap alatt 23.060 fővel nőtt az új fertőzöttek száma is. Ez a növekedési tényező 10,9-es értékét jelentette akkor. Ez nyilvánvalóan a fertőzöttek számát meghatározó procedura inhomogenitására utal. Vagy a korábban regisztrált, de a hivatalosan nem publikált eseteket ekkor adták hozzá utólag a fertőzöttek hivatalos számához, vagy erre a napra vonatkozóan jelentősen megváltoztatták a számbavételi procedúrát. Persze az lenne jó és elvárt, ha egy állami intézményrendszer a járvány kezdetétől ugyanazzal az eljárásrenddel venné számításba a fertőzöttek számát. Szóval, ha a fertőzöttek száma alakulásának vizsgálatánál el lehetne tekinteni a számbavételi procedura (ezen belül a tesztelés feltételeinek) változásaitól, inhomogenitásától. Különböznék az új fertőzöttek hivatalos száma inkább fogja mutatni e procedura változását, mind magát a járvány lefolyását és a lakosság átfertőződésének lehetséges arányát.

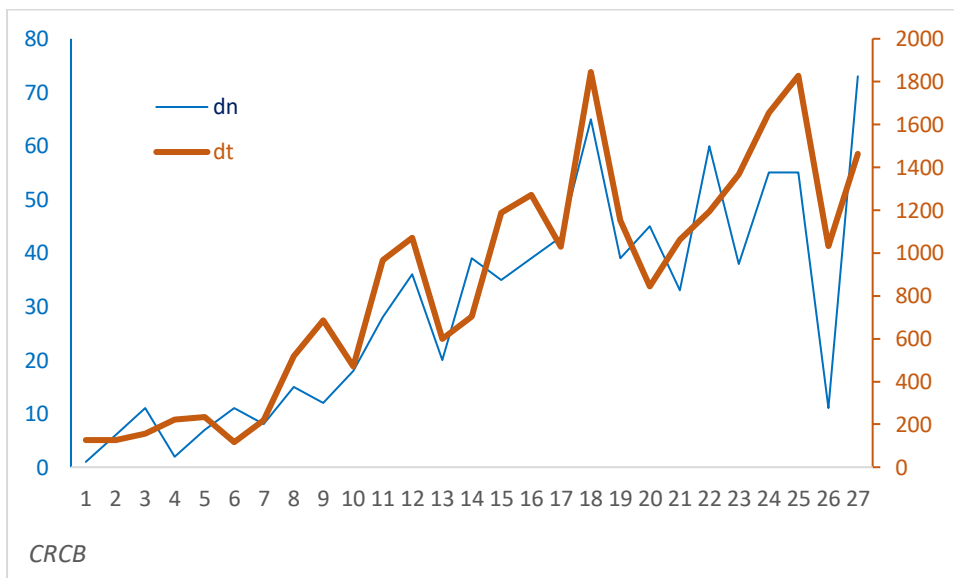
Az elvégzett tesztek száma és a fertőzöttek hivatalosan feltárt száma között nyilvánvalóan erős pozitív kapcsolat állhat fenn: minél több tesztet végez egy ország annál több fertőzött esetet tud feltárni. Persze az is érvényesülhet, hogy minél több esetet tárt már fel, és ha az adatott kormány szeretné elejét venni a járvány terjedésének, annál inkább igyekszik növelni a tesztek számát is. Ez következik abból is, hogy egy feltárt fertőzött eset társas kapcsolatainak nyomon követésére és tesztelésére van szükség ahhoz, hogy az esetlegesen fertőzött, de még tüneteket nem mutatókat is el lehessen különíteni. A két tényező közötti kétirányú kapcsolat feltárására több adatra és részletes idősoros elemzésre lenne szükség.

A két idősor közötti kapcsolat² nyilvánvalóan látszik a 3. ábrából. A fertőzöttek hivatalos száma és az elvégzett tesztek száma között erős pozitív korreláció van ($r=0,999$). Ugyancsak erős pozitív korreláció látható az új fertőzöttek hivatalos száma és az aznap elvégzett tesztek száma között is ($r=0,879$). Lásd továbbá a 4. ábrát.

² A tesztek napi számára vonatkozó adat forrása <https://444.hu/2020/04/07/magyarorszag-tovabbra-sem-erolteti-a-tesztelest>



3. Ábra: A fertőzöttek hivatalosan feltárt száma és a tesztek száma Magyarországon március 10-től.



4. Ábra: Az új fertőzöttek hivatalosan feltárt száma (dn) és az adott nap elvégzett tesztek száma (dt) Magyarországon március 10-től.

Legyen egy országban a feltárási ráta (δ) egy t időpontban

$$\delta_t = N_{o,t} / N_{a,t}$$

ahol N_o a fertőzöttek hivatalosan regisztrált száma, N_a a fertőzöttek tényleges száma.

$N_{o,t} < N_{a,t}$ minden t időpontra, és legyen $0 < \delta_t < 1$.

Legyen a fertőzöttek számának növekedési rátája (α) egy t időpontban

$$\alpha_t = N_{a,t} / N_{a,t-1}$$

Továbbá legyen a feltárási ráta változása (*inhomogenitási mutató*) a kormányzat által alkalmazott számbavételi procedúra inhomogenitása/instabilitása miatt $t-1$ időpontról t időpontra γ_t :

$$\gamma_t = \delta_t / \delta'_{t-1}$$

Ekkor a feltárt esetek számának változását úgy határozhatjuk meg, vesszük a t napra jellemző feltárási rátát és ezt elosztjuk a $t-1$ napig jellemző feltárási rátával (δ'_{t-1}). Ezt úgy is értelmezhetjük, mint egy inhomogenitási mutató- t , ahol az 1, vagy egyhez közeli érték a homogenitást, illetve kismértékű inhomogenitást jeleznek, és az ennél kisebb vagy nagyobb értékek pedig nagyobb eltérést a feltárási ráta t napig jellemző értékéhez képest. Ezek után az új fertőzöttek számának hivatalosan regisztrált számát megkapatjuk, a feltárási ráta t napon megfigyelt változása, a feltárási ráta, a fertőzöttek tényleges számának növekedési üteme és a fertőzöttek $t-1$ időpontban adott tényleges számának szorzataként:

$$\Delta N_{o,t} = \gamma_t * \delta_t * \alpha_t * N_{a,t-1}$$

Ennek megfelelően egy t napon az új fertőzöttek hivatalos száma függ attól, hogy *mennyire változott az adott országra jellemző feltárási ráta* t napon; volt e változás a fertőzöttek számát meghatározó procedúrában, vagy úgy is fogalmazhatunk, hogy *mennyire stabil az adott ország intézményrendszere által alkalmazott eljárásrend, mennyire képes homogén módon mérni*. Aztán függ magától a *feltárási ráta hosszú távon érvényesülő nagyságától*: mennyire volt képes t napig az adott ország intézményrendszere feltárni a járvány tényleges alakulását. Függ továbbá attól, hogy *a tényleges fertőzöttek száma* milyen arányban növekedett $t-1$ és t időpontok között, aztán természetesen attól, hogy *mennyi is volt ténylegesen a fertőzöttek száma* az adott országban a $t-1$ időpontban.

Lehetséges tehát, hogy egy ország kormánya a következőképpen dönt: szeretné, hogy úgy tűnjön, hogy megállította a járványt, ezért tudatosan korlátozni kezdi az elvégzendő tesztek számát. *Tehát tudatosan inhomogénná teszi a fertőzöttek számára vonatkozó procedúrát*. Csupán ebből a döntéséből az a következmény adódik, hogy csökken a fertőzöttek feltárási rátája ($\gamma < 1$). A kormány e döntése után egy külső szemlélő számára úgy fog tűnni, hogy csökkenni kezdett az új fertőzöttek hivatalos számának növekedése ($\Delta N_{o,t} / \Delta N_{o,t-1} < 1$) és ebből helytelenül azt a következtetést vonja majd le, hogy az új fertőzöttek tényleges számának növekedése is csökkent ($\Delta N_{a,t} / \Delta N_{a,t-1} < 1$). Tehát a kormány járványellenes fellépése sikeres volt, elérte a célját, megvédte az ország lakosságát a járvány pusztításától. Ez azonban hamis kép. A hamisság hosszabb-rövidebb idő múltán, de nyilvánvaló lesz. A járvány elsöprő erővel fog utat törni magának, úgy, hogy ez már a halálozási statisztikákban is egyre inkább látható lesz. Hiába próbáljuk rászorítani a fedőt a forró vizes lábosra, miközben úgy döntünk, hogy nem kapcsoljuk le a gázt – a lábosban lévő gőz ki fog törni. A legjobb esetben lerepítve a fedőt, vagy ha túlságosan erősek és akaratosak vagyunk, szétrobbantva a lábost. Ha figyeltünk volna a lábosban forró vízre, és kiengedtük volna a gőzt előbb, minden bajt el lehetett volna kerülni.

A csalásfelderítés nemzetközi irodalma (például [itt](#), [itt](#) és [itt](#)) az elemzett adatsorok (pl. adóbevallások, leadott rendelések, közbeszerzések, kártyajátékok, stb.) vizsgálatánál nagy figyelmet szentel a kiugró értékek (*outliers*) vizsgálatának. Az olyan értékeknek, amelyek egy adott eseménysor megoszlását figyelembe véve e megoszlás alsó és felső végén található, vagy másképpen olyan megfigyelési értékek, amelyek valamilyen szempont szerint szignifikánsak különböznek a többi megfigyeléstől (lásd [itt](#)). Ezek a megoszlások egy-egy értéke valamilyen cselekvő (magánszemély, cég, intézmény,

kártyajátékos, stb.) valamilyen aktivitásának köszönhető. Egy *outlier* lehet egyszerű véletlen eredménye, de létrejöhet az egyébként megfigyelt folyamat valamilyen torzításának eredményeként is. Ha az outlier-ek előfordulása rendszeresen egy-egy cselekvőhöz köthető (például a közbeszerzések esetében mindig X, a kormányfővel baráti kapcsolatokat ápoló „vállalkozó” cégcsoportja nyeri a legnagyobb tendereket), akkor annak valószínűsége, hogy tisztán véletlen hatásokkal állunk szemben, elhanyagolhatóan kicsi.

A fertőzöttek számát meghatározó - az állami intézmények közreműködését igénylő - procedura végeredményeként jön létre az új fertőzöttek hivatalos száma. Ezt persze mozgatja az is, hogy a fertőzöttek közül hányan és milyen gyorsan jutnak el oda, illetve tudnak eljutni oda, hogy a hivatalos intézményrendszer megállapítsa róluk, hogy fertőzöttek. Nehezen elképzelhető ezek után, hogy ha egy ország statisztikáiból rendre hiányoznak az outlierok, míg más országok járványügyi statisztikáiban meg rendre megjelennek – akkor e megfigyelés kizárólag a fertőzöttek magatartásához lenne köthető. Ez a jelenség tehát megfigyelhető, és empirikusan eldönthető az a kérdés, hogy az outlierok léte szignifikánsan gyakrabban fordul-e elő egy országnál, mint egy másikonál. Ha ezek előfordulása szignifikánsan magasabb egy országnál, akkor ez az eredmény az adott országban a fertőzöttek számbavételi procedúrájának inhomogenitására, és a mögötte meghúzódó anomáliákra utal.

A fenti kiválasztott országokban, az $N > 100$ esetek figyelembe vételénél az április 6-i és április 7-i furcsa értékek előfordulása nem kiugró gyakoriságú. Ha járvány kezdetétől eltelt időt, a fertőzöttek hivatalos számát is figyelembe vesszük és a felső és alsó 5%-ba tartozó értékeket számítjuk kiugró értéknek, akkor ezek (a 0,2 és a másnapi 6,64-es növekedési tényezők) előfordulása nem gyakoribb, mint a mintabeli más országoknál. A kiugró értékek előfordulásának esélye alapvetően a fertőzöttek hivatalos számával áll fordított összefüggésben (4. táblázat).

```
Logistic regression                                Number of obs   =      478
                                                    LR chi2(3)      =       9.19
                                                    Prob > chi2     =      0.0269
Log likelihood = -151.23419                       Pseudo R2      =      0.0295
```

	outl_5	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
test		.5908995	.4693469	-0.66	0.508	.1245692 2.802957
time		1.027047	.0268058	1.02	0.307	.9758301 1.080953
lnn		.7016151	.1027867	-2.42	0.016	.5264993 .9349752
_cons		1.014098	.757781	0.02	0.985	.2344353 4.386692

Magyarázat: test[1]: Magyarország, test[0]: egyéb országok; time: a napok sorszáma, attól a naptól fogva, amikor az adott országban $N > 100$ lett; lnn: a fertőzöttek hivatalos számának logaritmus; outl_5[1]: ha a növekedési tényező értéke az adott napon az alsó és felső 5% által kijelölt intervallumba esett.

4. táblázat: Outlierek előfordulásának becslése, logit, esélyhányadosok

Természetesen amennyiben az ilyen esetek gyakrabban fordulnának elő a jövőben például Magyarországon, akkor ez az eredmény is módosulhat. Mindehhez hozzá kell tennünk, hogy a kiugró értékek vizsgálatán túl a csalásfelderítés szokásos vizsgálati módszerei (pl. a Benford törvény) módok adnak másféle anomáliák vizsgálatára is.