

Adateltitkolás, ökonometria és a kommunista diktatúra

Nem várt helyről jött a segítség annak a kérdésnek a megválaszolására, hogy vajon milyen érvényességgel (*validity*) bír a CRCB által számolt J mutató a járvány erősségének alakulásáról. A J mutatót a CRCB az NNK (Nemzeti Népegészségügyi Központ) által közzétett, a szennyvízben mért koronavírus örökítőanyagának koncentrációjára vonatkozó “lakossági tájékoztatás” alapján számolja ki.

A “lakossági tájékoztatás” a koncentrációra vonatkozóan: “alacsony/mérsékelt/emelkedett/magas” szavakból válogat, a tendenciára vonatkozóan meg: “csökkenő/stagnál/emelkedik/erősen emelkedik” szavakat közli a magyar lakossággal rendszerint az ország huszonnégy megfigyelőpontjára (Budapest, Budapest agglomeráció és nagyobb városok) vonatkozóan. Lásd például [itt](#).

A szennyvízben a koronavírus örökítőanyagának koncentrációra vonatkozó számszerű adatokat tehát nem, csak a rájuk vonatkozó egyszerű szöveges értékeléseket közli a magyar kormány a magyar lakossággal. Az lehetne gondolni, hogy a hétről hétre közölt szöveges értékelések alapján nem lehet majd számításokat végezni, pontos képet kapni arról, hogy hol is tartunk a járványban jelenleg és mi is várható a következő hetekben. Pedig lehet.

Erre a CRCB, a konjunktúra kutatásokban ősidők óta, és széles körűen használt módszer nyomán adott példát a J (járvány) mutató megalkotásával. [A módszert az empirikus konjunktúra kutatásokkal foglalkozók “balance statistics”-nak hívják, szinte minden mutatójukat e módszerrel számolják; a módszert [Henri Theil](#) ajánlotta még az 50-es évek végén, a konjunktúra kutatások hőskorában; [itt](#) lehet még olvasni róla. Ezzel a módszerrel tehát a kvalitatív (egyébként ordinális) szöveges értékeléseket, mint pl. “rossz / kielégítő / jó”, illetve az ezekből képzett idősorokat fel lehet használni ökonometriai elemzéseknél, a gazdasági konjunktúra előrejelzésénél.] A CRCB módszere [egyszerű](#): feleltessünk meg az “alacsony/mérsékelt/emelkedett/magas” szöveges értékeléseknek számokat: “1/2/3/4” úgy, hogy az “alacsony” kapjon 1-es értéket (ha “alacsony” akkor 1) a “mérsékelt” kapjon 2-es értéket (ha “magas” akkor 2), stb. Járjunk el így az örökítő anyag koncentrációjának és tendenciájának esetében is. Aztán az egyes városokra kapott értékeket adjuk össze – és meg is van a J mutató, ami a járvány országos alakulását mutatja. [Az egyes városi adatok lakosságarányos, vagy a szennyvíz mennyiségével meghatározott súlyozásával nem foglalkoztunk.] E módszer alapján számított idősorokat 2020. november 16. óta közöljük [a CRCB honlapján](#).

Az ökonometria, és ezen belül az empirikus konjunktúra kutatásokban használt módszerek ismerete tehát jó szolgálatot tehet, amikor adateltitkolással találjuk szembe magunkat. Van már idősorok, amely a járvány országos lefutásáról tudósít – annak ellenére, hogy semmilyen pontos mérési adat nem áll rendelkezésünkre.

De mindezidáig nem tudtuk, hogy mennyire jó, vagy rossz az amit a J mutatóban a CRCB számított.

Március 17-én azonban az NNK publikálta a “Koronavírus monitorozás szennyvízben: országos COVID-19 korai előrejelző rendszer” című [kutatási jelentését](#), amely már közelebb visz e kérdés megválaszolásához.

E jelentés 12. oldalán szereplő 2. ábra [“Országos szennyvíz adatok lakosságárányos súlyozott átlaga (2020 június-2021 február)”] a fontos most számunkra. Ez tartalmazza a koronavírus örökítőanyagának koncentrációjára vonatkozó országos (lakosságárányal súlyozott) eredményeket.

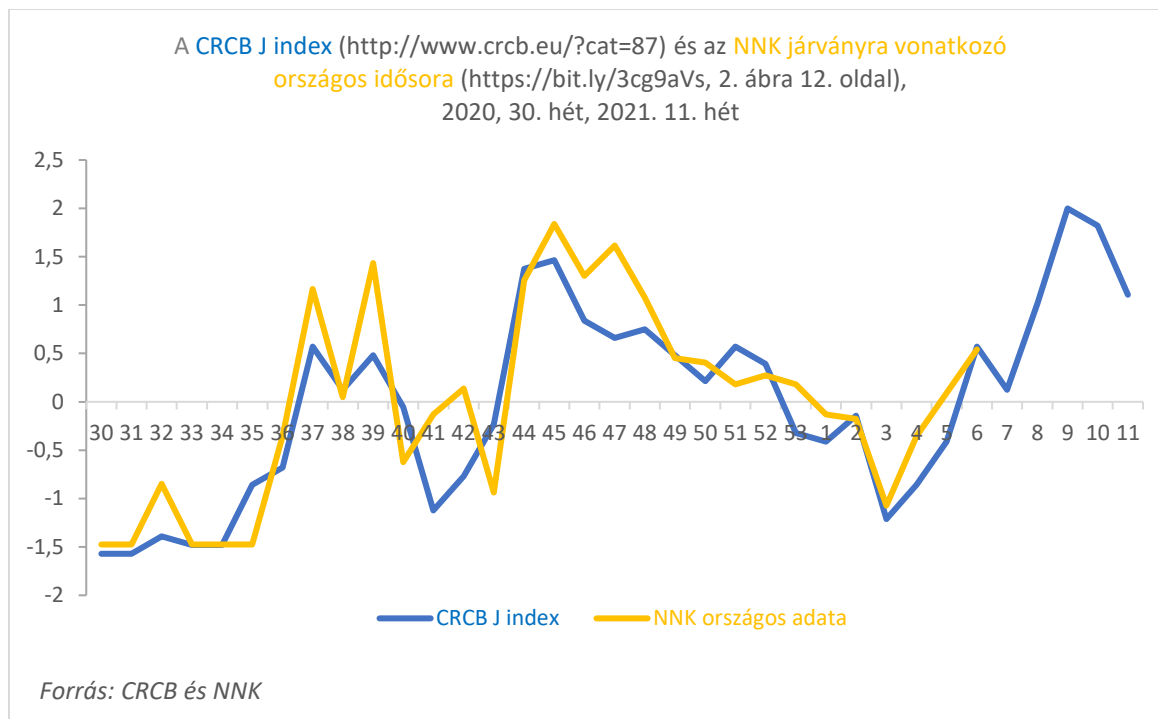
Az első nehézség ezzel azonban mindjárt az, hogy az ábrán szereplő adatok nem találhatóak meg a kutatási jelentésben. Bár a jelentés ugyanitt megállapítja: “A járványhelyzet országos értékelése céljából a szennyvízminták fekális Enterococcusra normalizált SARS-CoV-2 koncentrációiból hetente lakosságszámmal súlyozott átlagot számolnak (2. ábra). A heti adatokról és különösen a szokatlan változásokról tájékoztatás készül az országos tisztifőorvos és az Operatív Törzs számára.”, de az ábrán szereplő adatok nem szerepelnek a tanulmányban.

Ekkor jön segítségünkre az, hogy a CRCB kutatóinak egy része jól ismeri a kádári ‘80-as évek klasszikus módszerét, az ún. “centiméter módszer”-t. (Ez például némely, az adósságteherre vonatkozó belső pártjelentés ábráinak kislabizálásánál tett jó szolgálatot annak idején). A centiméter módszert ugyan a nemzetközi irodalomban jóval kevésbé használják, mint az előbb említett Theil féle egyenleg statisztikát, de ennek ismerete most fontos a továbblépéshez.

Nagyítsuk ki a közölt ábrát és a függőleges tengelyen mért értékeket mérjük le a számítógépünk monitorján centiméterben! A mérés közben elkövetett hibák nem számítanak, mivel ezek véletlenszerűek lesznek (ún. “fehér zaj”) és egészen kicsik, szóval e tévedések nem fognak számítani az elemzés során. Írjuk le aztán és rendezzük idősorba a kapott értékeket!

Az egyszerűség kedvéért sztenderdizáljuk a két idősort (számítsuk ki a z értékeket), hogy együtt könnyen ábrázolhatók legyenek – a függőleges tengelyen szereplő dimenzió úgyse számít annál az elemzésnél, hogy a két idősor mennyire mozog együtt.

Aztán hasonlítsuk össze a két idősort! [Az [ADF teszt](#) szerint a két idősor különbsége stacionárius – ami kiindulásként jó jelnek tekintünk.] A két idősor közötti korreláció értéke 0,89. A két idősor szorosan együtt mozog, ami a mellékelt ábrán is jól látszik.



Tehát a CRCB J mutatója egészen jól mutatja az NNK által az eredeti szennyvíz adatok alapján számolt országos tendenciát. A két időszor közötti különbség abból is adódhat, hogy míg a CRCB J mutatója a városok adatait súlyozatlanul összesíti, addig az NNK időszora úgy, hogy a településeken lakók számával súlyoz.

A CRCB J mutatója tehát segíti a magyar állampolgárok tájékoztatását: azt az információt, amit az országos tisztifőorvos, az “operatív törzs”, és feltehetően Orbán Viktor is megkap, immár láthatják a magyar állampolgárok is. Az előbbieket nem lesznek kivételezettek.

Ezen túl az NNK jelentésében szereplő 2. ábra mindössze a 6. hétig mutatja a SARS-CoV-2 örökítőanyaga koncentrációjának alakulását, és tudósít a járvány állásáról, addig a CRCB már közzétette a 11. heti adatokat is.

Forrás:

https://www.facebook.com/permalink.php?story_fbid=3878899335537264&id=175386655888569

Az elemzés alapadatai itt érhetők el: http://www.crcb.eu/wp-content/uploads/2021/03/covid_19_2020_crcb_data_210323.zip