

## 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.

Az magyar kormány az NNK által hétről-hétre közölt "lakossági tájékoztatás"-sal meg kívánja óvni a magyar lakosságot attól, hogy a pontos, számszerű adatok nyilvánosságra kerüljenek, és ezért csak egyszerű, szöveges értékeléseket közöl. Ezzel a magyar kormány igazodik a "magyar lakosság" általa vélt értelmi szintjéhez: számokat a magyar emberek úgysem fognak megérteni.

Tudjuk, hogy ez nem az NNK szakértőinek hibája, nem az NNK szakértői akarták így.

Ezért a "lakossági tájékoztatás" a koncentrációra vonatkozóan: "alacsony/mérsékelt/emelkedett/magas" szavakból válogat, a tendenciára vonatkozóan meg: "csökkenő/stagnál/emelkedik/erősen emelkedik" szavakat adja meg rendszerint az ország huszonegy megfigyelőpontjára (Budapest 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. Olyan ez, mintha az Országos Meteorológiai Szolgálat minden nap az egyes városokra jellemző pontos hőmérséklet adatok megadása helyett csak ennyit mondana "nagyon csúnya idő van / csúnya idő van/ szép idő van / nagyon szép idő van" és "ma rosszabb lesz az idő mint tegnap / ma ugyanolyan az idő lesz mint tegnap / ma jobb lesz az idő mint tegnap".

A magyar kormány talán arra számít, 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 de.

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, akit érdekel. Ezzel a módszerrel tehát a kvalitatív (egyébként ordinális) szöveges értékeléseket, mint pl. "rossz / kielégítő / jó", illetve az ezekből képzett idősorokat fel lehet használni ökonometria 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 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 ismeretete tehát jó szolgálatot tehet, amikor adateltitkolással találjuk szembe magunkat. Van már idősoruk, amely a járvány országos lefutásáról tudósít – annak ellenére, hogy semmilyen pontos mérési adat nem áll rendelkezésünkre.

De mindeztidáig 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 lakosságarányos súlyozott átlaga (2020 június-2021 február)"] a fontos most számunkra. Ez tartalmazza a koronavírus örökítőanyagának koncentrációjára vonatkozó országos (lakosságaránnyal súlyozott) eredményeket.

Az első nehézség ezzel azonban mindjárt az, hogy az ábrán szereplő adatok nem találhatók 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 adatokat nem köti a magyar állampolgárok orrára, ezek nem szerepelnek a tanulmányban. Ismét vannak tehát egyenlők és egyenlőbbek.

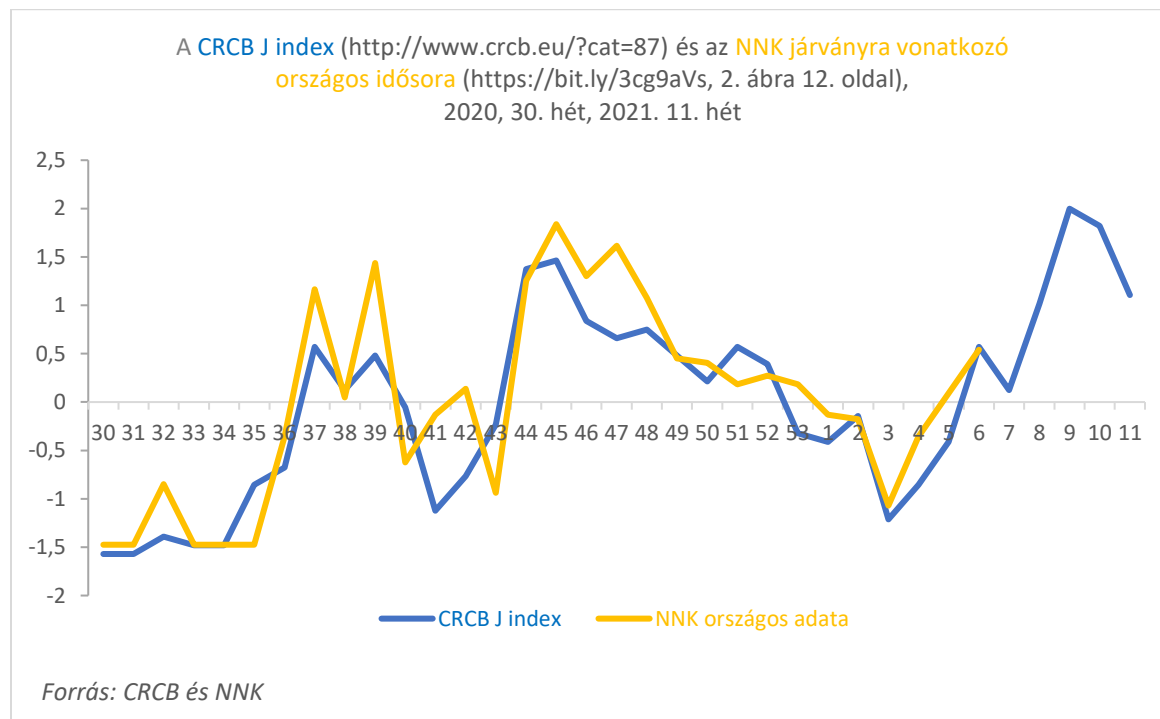
Egy nyugat-európai kutató ekkor talán leragadna és befejezné a munkát. De nem azok, akik egy kommunista rendszerben szocializálódtak. Ekkor jön segítségünkre az, hogy a CRCB kutatóinak egy része jól ismeri a kádári ‘80-as évek klasszikus módszerét, az ún. “centiméter módszer”-t. (Ez például némely, az adósságteherre vonatkozó belső pártjelentés ábráinak kisilabizálásánál 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 egyre fontosabbá válik a NER vezérelt Magyarországon.

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 mindkét idősor stacionárius és természetesen különbségük is 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 idősor közötti különbség abból is adódhat, hogy míg a CRCB J mutatója a városok adatait súlyozatlanul összesíti, addig az NNK idősora 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.

Az adatok eltitkolására vonatkozó kormányzati törekvés tehát kudarcot vallott: nem számolt sem az ökonometriaival módszerekkel, sem azzal, hogy az adateltitkolás ténye déjà vu élményt fog keltetni. Emlékeztetni fogja a kutatókat arra, hogy ilyent a kommunista diktatúrában láttak már.

A szöveges értékelés alkalmazása ("ma jobb, mint tegnap") nem megoldás. Az eredmények nem közlése, csak ábrán való "bemutatása" nem megoldás.

E bejegyzés végén a szakpolitikai tanulságokat egy mondatban is össze lehet foglalni: az adatok eltitkolására vonatkozó kormányzati politika csak akkor lehet sikeres, ha a magyar kormány a járványról semmilyen adatot nem tesz közzé a jövőben.

Forrás:

[https://www.facebook.com/permalink.php?story\\_fbid=3878899335537264&id=175386655888569](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](http://www.crcb.eu/wp-content/uploads/2021/03/covid_19_2020_crcb_data_210323.zip)