Bizonyítás-szimulátor, avagy mit tudhat ténylegesen egy big-data alapon eljáró ROBOT-COVID-szakértő?

(Evidence-simulator, or what can be derived based on big-data by a ROBOT-COVID-expert?)

Pitlik László, MY-X team

Kivonat: A bizonyítás-szimulátor egy olyan komplex modell-sorozat, mely a mindenkor ideálisnak vélhető adatvagyon (big-data) alapján azt az elemzési potenciált hivatott feltárni, tesztelni, minősíteni, mely a vizsgált jelenséghalmaz esetén releváns kérdésekre automatizált válaszokat képes levezetni. Az oktatás kapcsán (is) triviális kerettörténet lehet a COVID19 pandémia, s ennek idealizált adatvagyona egy 100-fős szigetet feltételezve, ahol a betegséget pl. hajótörött denevér indítja el a sziget egy adott pontján, ahol a szigetről (ennek környezeti tényezőiről), a szigetlakókról (ezek életvezetéséről, mikrobiális múltjáról – vö. old-friends-logika), ill. a személyes digitális nyomkövető berendezések által minden egyes szigetlakó útvonalairól, tartózkodási helyeiről, találkozásairól, ill. rokonsági fokairól, együttélési paramétereiről, stb. minden elvárható adat az elvárható pontossággal/minőségben rendelkezésre áll. Az ideális adatvagyon speciális aspektusa a bármilyen tesztek létezése, vagy teljes hiánya. Az így, azaz napról napra (pillanatról pillanatra) gyarapodó „big-data” erőtér alapján a robot járványügyi szakértő az alábbi kérdésekre keresi a válaszokat: Ki fertőzött meg kit, mikor? Elegendő-e egy találkozás ez fertőzéshez? Mennyire hosszú idejű találkozásra van szükség egy fertőzéshez? Milyen időbeli távolság számít adott helyen járó két személy esetén még fertőzési lehetőségnek? A fertőzött személy mikortól fertőz? Van-e a megfertőzhetőség szempontjából bármilyen életforma/kor/nem/egyéb-erőtérkülönbség egyed és egyed között? Milyen gyorsan lehet (ha egyáltalán) újra fertőződni, stb.?! Az adatvagyon értelemszerűen véletlenszám-generátorral kerül kialakításra és a cél az adatvagyon adott állapotát értelmezni képes modellek kialakítása, de még inkább a modelleredmények értelmezését végző hermeneutikai/döntéstámogatási alrendszer minél részletgazdagabb levezetése. A keretrendszer alkalmas arra, hogy informatikai képzések tartós, valóságközeli, Hallgatók és oktatók számára is állandó és megújuló motivációt jelentő keretrendszereként hasson. Valós adatvagyon esetén a rendszerfejlesztés már kellő minőségi jegyekkel rendelkező változata azonnal hadra fogható kell, hogy legyen. A keretrendszer a rendszertervezés, rendszertesztelés, minőségmenedzsment, rendszerüzemeltetés, stb. alapfogalmak kapcsán tételes játszóteret nyújt a KNUTH-i elv megvalósulásának: tudás az, ami forráskódba, jelen esetben egy ROBOT-COVID-szakértőbe átírható, minden más emberi aktivitás (a konkrét téma kapcsán hangsúlyozottan) CSAK művészet. A keretrendszer minden egyes részeredménye min. egy-egy TDK-dolgozat alapjaként értelmezhető, maga a keretrendszer sikeres kialakítása illik, hogy egy PhD-címet alapozzon meg. Minden köztes állapot BSC/BPROF, ill. MSC szintű informatikai (egyéb pl. biológiai) kihívásokat kell, hogy tartalmazzon. A keretrendszer lépéseinek egyedi komplexitása nem illik, hogy meghaladja az ECDL szintet. Az így kialakított szimulátor minőségbiztosítása is triviális: egyrészt választ lehet adni arra a kérdésre: kit érdemes leginkább adott pillanatban tesztelni. Amennyiben pedig a tesztelés megtörténik: akkor objektíven látható, bevált-e a gyanú? Azért érdemes majd csatlakoznia önként mindenkinek a rendszerhez, mert bárki közelít bárkihez, pontosabban senki nem lehet és nem lesz képes feltárni a találkozás kockázatát, mint a szimulátor alapján. Vagyis nem utólag kell jelezni, lehetett-e probléma, hanem előre – preventív jelleggel. Így számos kontaktus nem kell, hogy elkerülésre kerüljön, csak azok, melyek tényleg kockázatosak. Persze egyetlen egy becslés sem lesz teljes mértékben hibátlan. Ez lehet furcsa személyes döntési erőterek hajtóereje…

Kulcsszavak: hasonlóságelemzés, pivot/OLAP-riportok, fertőzések forrása, kontakt-kutatás, fertőzőerő, újra-fertőződés

Abstract: The evidence simulator is a complex model-series for relevant questions where the maximum of the analytical potential should always be derived which is realistic based on the best big-data resources. This analytical potential may not use any human influence – each of their steps should be done in an automated way. The education (especially in IT-based cases) needs a frame where the creation-oriented teaching and learning ensure the most relevant level of motivation of Students. The COVID as such could be one of these frames. The evidence simulator could be developed without the idealized data asset. It is possible to derive such a database in a semi-randomized way where the real data might be identified in the future. The storyline could be: 100 persons come to an island where an ill bat will be responsible for the first infection(s). The ideal data asset should contain data about the geography, climate, etc. of the island, parallel to the data about the 100 persons (incl. microbiological old-friends, genetical information, etc.) and first of all about the movements of each individual on the island (based on GPS tracking). It is a specific information when we know about test-results (c.f. being infected, being after an infection, etc.). The big-data will be step by step a real big-data. We (human beings) have only to define questions for the robot expert like: Is enough one single contact for an infection? Who could be responsible for the infection of a given person? How long should be 2 persons in contact and within which distance? Is it possible (at all) a re-infection? From what time to what time is a given infected person danger for (which type of) other persons? Etc. The analytical process should be able to say: I am sorry, I do not know (here and now). This kind of frame, this kind of challenge could involve quasi each aspect of an IT-oriented education where the Students should be able to develop and maintain the hole system incl. quality management. This challenge could demonstrate what means the KNUTH-principle. This challenge could deliver useful fragments/systems for the society instead of irrational data and qualitative low analyses with a lot of human influence in the background. There is one single question concerning this framework: the GDPR!

Keywords: similarity analysis, pivot/OLAP-reporting, source of infections, contact tracing, infectability, re-infection

# Bevezetés

Az alábbi hivatkozások a hasonlóságelemzésekre alapozó egészségügyi jelenségek esettanulmányait mutatják be kiragadott példákon keresztül:

Előzmények:

* <https://miau.my-x.hu/miau2009/index.php3?x=e143> (A circaseptan ritmus a COVID-19 megbetegedések és halálesetek kapcsán az adatvagyon-gazdálkodás minőségbiztosítása keretében)
* <https://miau.my-x.hu/miau2009/index.php3?x=e0&string=cs%C3%ADp> (Csípőprotézis műtétek hozzáférhetősége kistérségi adatok tükrében)
* <http://miau.my-x.hu/miau/242/perinatal_risk_models_4_szerzos.docx> (Perinatális kockázatok modellezése a Tauffer-statisztika keretében innovációs folyamatok támogatására)
* <https://miau.my-x.hu/miau/112/tdk/gygyszr/> (A hasonlóságelemzés lehetőségei a gyógyszerek ár-kockázat (mellékhatás) viszonyának meghatározásában)
* <http://miau.my-x.hu/miau/106/gyogyszer_v2.doc> (A gyógyszer-választás közgazdasági és egészségpolitikai vonatkozásai)
* <http://miau.my-x.hu/miau/187/pharma_v2.doc> + <http://miau.my-x.hu/miau/187/pharma_v1.doc> (A magyar gyógyszeripar - hagyományos szemmel + A magyar gyógyszeripar - elemzői szemmel)
* <https://miau.my-x.hu/mediawiki/index.php/Gy%C3%B3gyszerek> (Ár/teljesítmény-elemzés gyógyszerek esetén)

# Egy elemzési típushelyzet bemutatása

Kivonat egyben egy fajta rövid kutatási terv, mely annak alapjait rakja le, miként lehet pl. IT-képzések keretei között olyan társadalmilag racionális célokat kitűzni, melyek mentén az alkotás-orientált oktatás és jegyadás/jegyszerzés vélelmezhetően motivált Hallgatóságot generál.

Az alábbi esettanulmány a jelen pillanatban (2020. decemberében) legaktuálisabb kérdések egyikét feszegeti: minden (a későbbiekben már majd csak COVID) oltóanyag tekinthető-e másként egyformán kockázatosnak – nem konkrét páciens, hanem úm. az átlagember esetében?

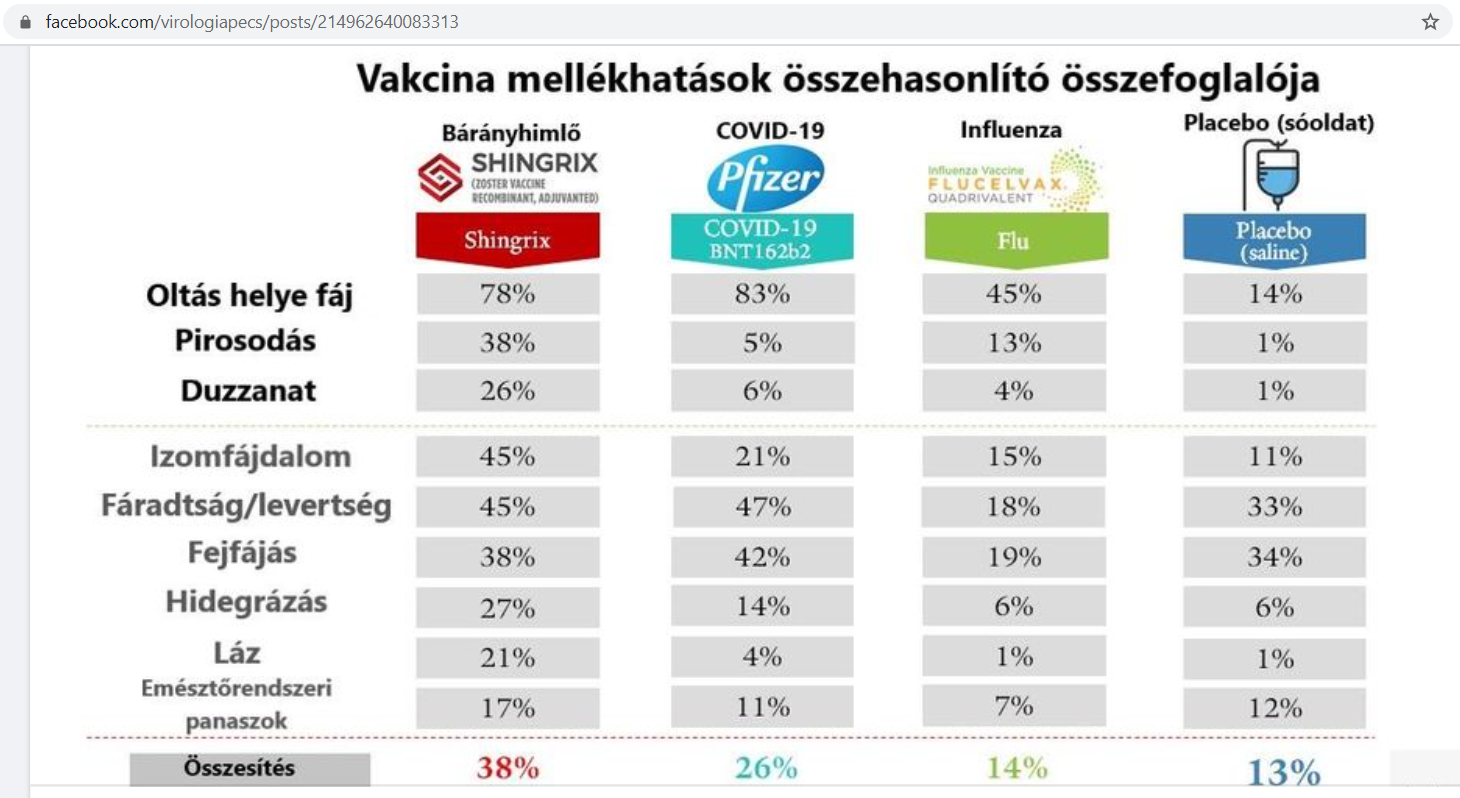
Hasonlóan releváns, de egyelőre adatok nélkül elemzésre alkalmatlan kérdések lehetnek pl.

* Fertőz-e egyáltalán egy oltott páciens? S ha igen, mettől meddig?
* Mely oltóanyag egyedenként kinek mekkora/milyen kockázatot jelent (genetikai és/vagy hagyományos leíró adatokra támaszkodva)?

Az alábbi táblázat a közösségi média és a TV-adások kapcsán elvileg közismert. A típushelyzet semmiben nem különbözik az előzmények között látható gyógyszerválasztási problémától. Ismert egy (itt és most ráadásul nem a szerzők által számszerűsítendő) OAM, melynek objektumai az oltóanyagok (4 db), ill. mellékhatások (attribútumok – szám szerint: 9 – mert az összesítés lenne éppen az a megkérdőjelezett marketing üzenet, mely alapján a naiv szemlélő a covid oltóanyag itt leírt alakzatát közepesen kockázatosnak illene, hogy érzékelje (38%>26%>14%). A kérdés már csak az: tetten érhető-e egy fajta hermeneutikai érzékcsalódás? (vö. <http://miau.my-x.hu/miau/237/VizMan%20ENG.docx>)

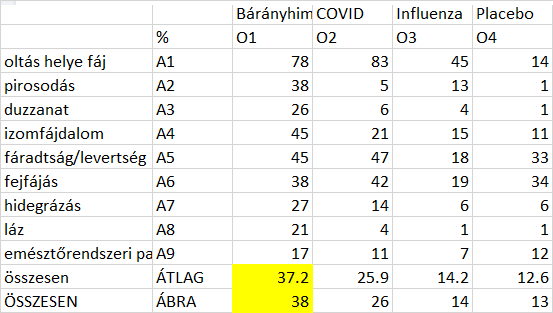
Az elemzési folyamat egyszerű: miután újra rögzítésre kerültek a számadatok, minden objektum kap egy norma-értéket (100 jóságpontot jelentő konstans-értéket), melyet hasonlóságelemzéssel igyekszünk közelíteni (vö. kiselemszámú elemzések pl. <https://akjournals.com/view/journals/032/70/2/article-p275.xml>)

Az elemző szolgáltatás elérhetősége: <https://miau.my-x.hu/myx-free/>, ill. <https://miau.my-x.hu/myx-free/coco/index.html>



1. Ábra: Az OAM (forrás: facebook)

További részletek: <https://miau.my-x.hu/digeco/2020/2020osz/covid_oltoanyag_mellekhatasok.xlsx>



1. Ábra: Az összesen sor ellenőrzése (forrás: saját ábrázolás)

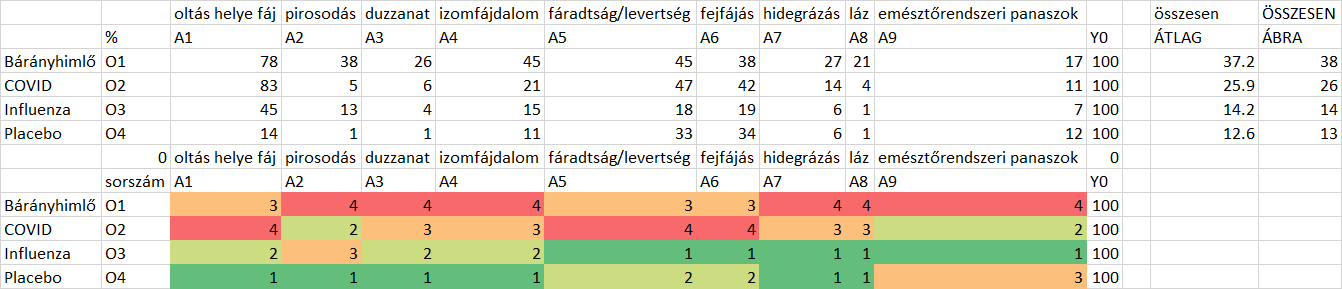
A 2. ábra alapján triviálisan látható (sárgával is kiemelve), hogy az adatközlők annyi fáradságot sem vettek, hogy a mellékhatásonként egész számra kerekítve megadott átlagos (összesen) százalékértéket összevessék a kerekítés eredményeként (vélhetően helyes alapszámokból helyesen levezetett) 38 %-os értékkel. Ilyen esetben, amikor a páciensek bizalma a tét, minden apró zavar az egészségügyi rendszer számára komoly kockázatot jelent (ha csak meggyőzési munkát/költséget is). Az is világos természetesen, hogy a kerekítés kapcsán csak véletlenszerűen jöhet ki a részek és a mindösszesen vizuális viszonya minden zavar nélkül. Mégis igaz azonban (pszichológiai fókusszal), hogy az egészértékű kerekítés +/-1 százalékos toleranciája ott, ahol egy adott táblázatban 1% a különbség két kezelés értéke között (vö. 14% vs. 13%), már eléri azt a zavaró szintet, ami ellen a tizedesjegyek megadásával kell és lehet fellépni. Ha a kerekítési zavar a 13-as és 14-es számok kapcsán, s nem a 38-as/37-es szám kapcsán lépett volna fel, a táblázat információértéke (rémhír-potenciálja, összeesküvés-elméletgenerálási potenciálja) sokkal nagyobbá vált volna – teljesen feleslegesen. Érdekes felvetésnek számít annak feltárása, vajon a sok tizedesjegyes közlés mennyivel veti vissza adott táblázat fogyasztását/olvasottságát – s mennyivel növeli a kevés tizedesjegy a fenti anomáliákon keresztül a bizalomvesztést/rémhír-terjesztési hajlandóságot? A

Az ideális adatközlés ott, ahol a 37.2<>38 semmilyen kerekítés alapján eset előállhat az, ha minden egész-értékű alapadat helyett több (de legalább még egy) tizedesjegyet IS megadnak – különösen akkor, ha a fenti +/-1 százalékos (matematikailag korrektnek számító) ingadozás a kezelések közötti különbségek 1%-os szintjét direktben érinti…

Hasonló (gondatlansági?) anomáliák már korábban is ismertek voltak: vö. pl. az ásványvizek iontartalmának összege csak esetlegesen volt azonos az összes iontartalommal egy úm. hivatalos forrás esetén (vö. <https://miau.my-x.hu/miau2009/index.php3?x=e0&string=%C3%A1sv%C3%A1ny>)

Szomorúan kell megállapítani, hogy a bulvár újságírás szintjén is érdekes és értelmezhető anomáliákat nem sikerült évtizedek alatt sem levetkőznie szakterületeknek?! Ezek az anomáliák és az ezek ellene lépések vajon miért nem képezik részét a magyar nyelv és a matematikai oktatásnak már általános iskolától kezdődően?

# Az elemzés



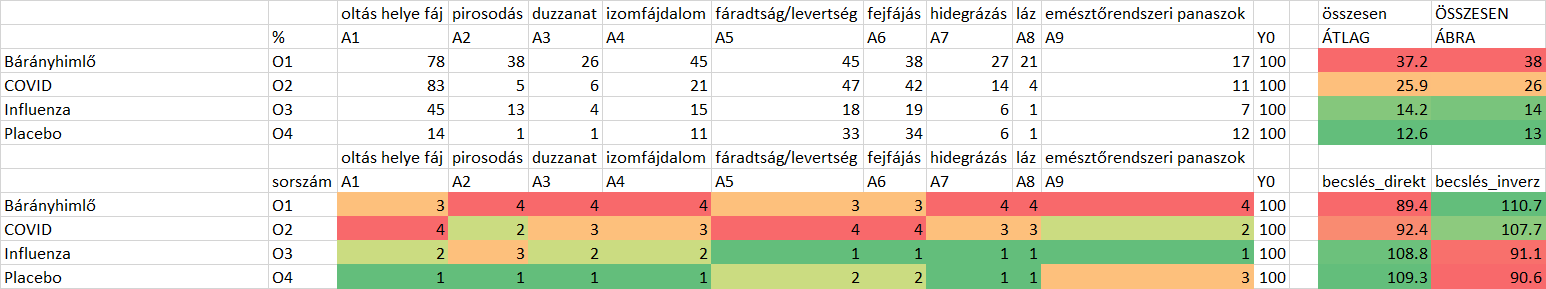
1. Ábra: A sorszám-nézet (forrás: saját ábrázolás)

A 3. ábra és ennek színkódjai az 1-2. ábra után még egyértelműbbé teszi, hogy a placebo, mint objektum nem minden esetben rendelkezik az 1. sorszámmal, azaz a legkisebb értékkel. Értelemszerűen minden mellékhatás értelmezési iránya triviálisnak tűnik elsőre: minél kisebb a mellékhatások fellépésének aránya, annál ideálisabb az oltóanyag!

Itt két dolgot is meg kell jegyezni:

* Az, hogy a placebo nem egy fajta minimum-jellegű benchmark, már ab ovo felveti, hogy olyan adatokra nincs is szükség (vö. kérdőívek), melyek az adatközlő szubjektív észrevételeit tartalmazzák. Egy ilyen fontos kérdésben csak a mérhető adat illene, hogy adat legyen – vagyis ezen százalékos értékek az oltást követő megfigyelés alapján kellene, hogy létrejöjjenek.
* Emellett a média által felkért orvosok egyike-másika szerint: nem baj az, ha az oltás helye úm. jelez! Tehát nem kizárt, hogy a fenti 9 attribútumból az 1-2-3 iránya nem is triviális? A 4. pedig az oltás esetleges rossz technikájú beadásának inkább a következménye, mint az oltásnak/oltóanyagnak magának?

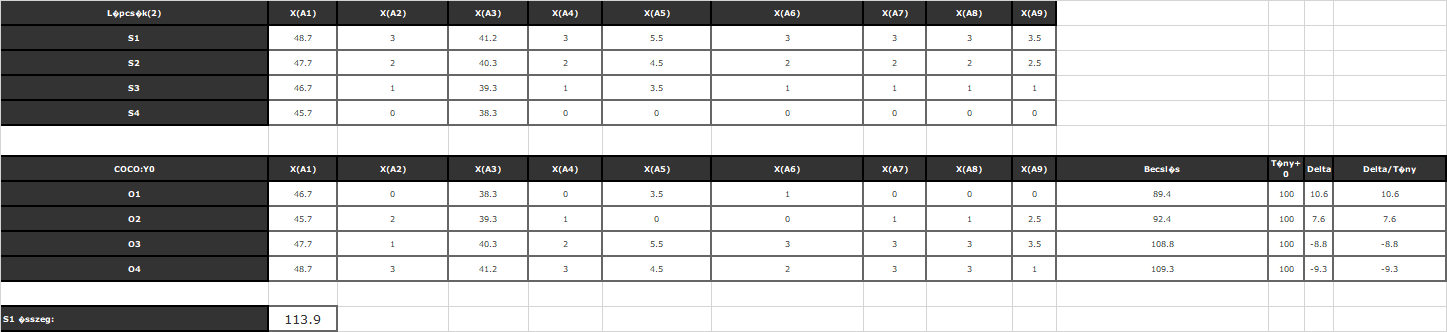
A 3. ábra alapján annak óvatos gyanúja felmerülhet a szemlélőben, hogy a naiv átlagolás által sugall objektum-sorrend a színkódok mozaikossága folytán nem feltétlenül lesz egyértelműen levezethető egy anti-diszkriminatív matematikai (optimalizáló, mesterséges intelligencia-alapú) eljárással, mert nincsenek vonalkód-jellegű színsávok (azaz csak zöld vs. csak piros cellák egy-egy sorban).



1. Ábra: Az eredmények (forrás: saját ábrázolás)

A 4. ábra alapján látható és belátható:

* A robot elemzése valid minden objektumra, mert a direkt és inverz modellek (melyek ellentétes inputokból indulnak – az elvárásoknak megfelelően ellentétes outputra is jutnak a 100-as normaértékhez képest).
* A placebo és az influenza-oltás között a távolság jelképes (109.3 vs. 108.8 jóságpont, ahogy a 13% és 14% is, szemben a 12.6 vs 14.2-es számpárral).
* A himlő- és a covid-oltás között a távolság nagyobb, mint a placebo és az influenza-oltás között az átlagoknak is megfelelően.
* Ami azonban a színkódokat illeti: a közölt összesen-színsávok 4-rétegűságe és a számított/becsült idealitások alapvetően inkább 2-rétegűsége üzenettartalmát vizsgálva, már karakteresen eltér egymástól.
* A modellszámítás nem talált olyan részleges attribútum-együttállást, mely alapján minden oltóanyag másként egyformának tűnne az adatok egy része alapján.



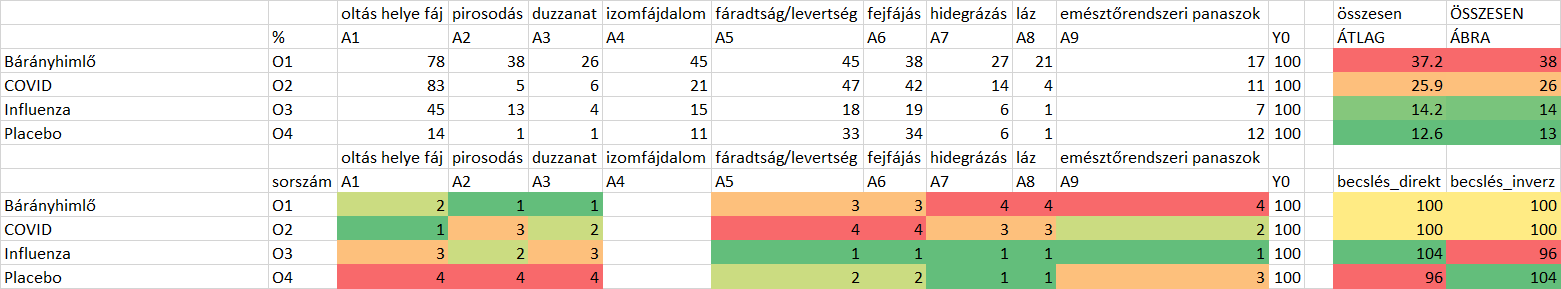
1. Ábra: Részletek (forrás: saját ábrázolás)

Az 5. ábra alapján vélelmezhető, hogy

* A 113.9-es (alul) S1-összeg, amikor is minden mellékhatás egyetlen egy objektum esetén a legjobb 4.6 jóságponttal kisebb, mint a placebo 109.3-as értéke.
* 4.6 pont nincs a 89.4 és a 92.4 között, vagyis a placebo-t érintő adattorzulás becsült mértéke nagyobb, mint a két gyengébb és egymástól távolibb (norma alatti) objektumpár távolsága.
* Az A2, A4, A6-7-8 attribútumok nem kellettek ahhoz, hogy a minden objektum másként egyforma elvet ki lehessen zárni.

# Alternatív elemzés

Tegyük fel, hogy az orvosi intuícióknak (vö. jelezzen az az oltás, ill. oltásbeadási hiba az izomfájdalom) van racionális alapja, s a bemeneti jelek közül a 4. kitörlésre kerül, s az első három (A1-2-3) pedig ellentétes irányt kap:

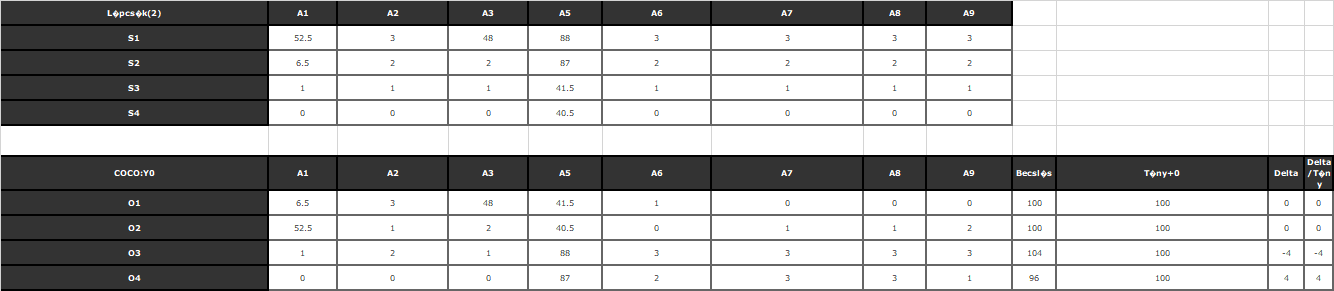


1. Ábra: Az orvosi intuíció OAM-ja (forrás: saját számítások)

A 6. ábra tehát bemutatja:

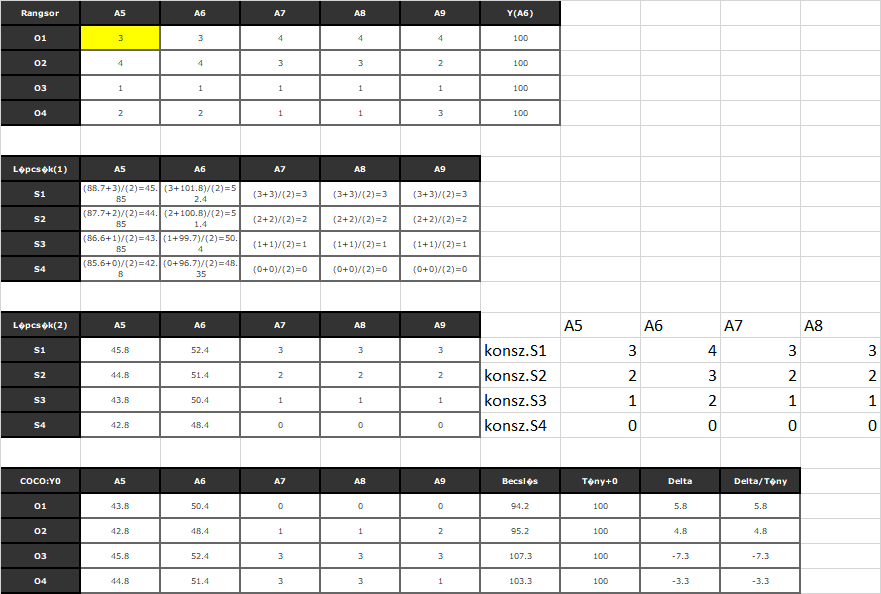
* A hiányzó A4 (helyét).
* A megváltozott A1-2-3 színkódjait.
* S a megváltozott eredményt, mely értelmében: a placebo a legrosszabb helyre került.

Ez a vélelem és a 7. ábra A6-7-8-9 fel nem dolgozottsági jelei (vö. kötelező minimum a lépcsős függvényben 0-1-2-3 4 objektum esetén) arra a következtetésre kényszerítenek, hogy csak az A5-6-7-8-9 attribútumokkal is készüljön egy elemzés.



1. Ábra: Részletek (forrás: saját ábrázolás)

# A5-6-7-8-9



1. Ábra: Az A5-6-7-8-9 hatása (forrás: saját számítások)

A 8. ábra alapján a következők vonhatók le:

* A placebo elvesztette ideális helyzetét, ami nonszensz – vagyis az adatvagyon értéke megkérdőjelezhető.
* A konszolidált lépcsőértékek csak az A6 esetében térnek el a kényszerpályától (0-1-2-3).

# Konklúziók

A 4-objektumos elemzések alacsony objektumszáma miatt matematikailag eleve kockázatosak – context free módon. Kontextus-függően ezt erősíti fel a placebo-adatvagyon esetlegességi gyanúja. Szerencsére magát az alapüzenetet, miszerint a covid-elleni oltás adott formája nem a legrosszabb objektum, nem kell elvetni, de a covid-elleni oltás nem triviálisan jobb, mint a himlő elleni oltás és mindkettő rosszabb, mint a másik két vizsgált alternatíva.

# Referenciák

…szövegközben…