Szubjektív egészségminősítési adatvagyonok elemzési értékének automatizált feltárása I.

(Automated derivation of the analytical value of subjective health assessment data assets)

Fejes Zsolt (0000-0001-9065-5358)\*

Pitlik László (0000-0001-5819-0319)\*

Szűcs Diána (0009-0004-7680-6948)\*

Rikk János (0000-0002-3846-6661)\*

Mészáros Melissza (0009-0006-1391-6220)\*

Gáncs Róbert (009-0003-6570-3129)\*

Lojek Bence (0009-0007-1643-868X)\*

Pitlik László, ifj. (0000-0002-8058-9577)\*

Pitlik Mátyás (0000-0002-1991-3008)\*

\* MY-X team

Kulcsszavak: konzisztencia, validáció, minőségbiztosítás, mesterséges intelligencia, tartalom-függetlenség, kérdőív

Keywords: consistency, validation, quality assurance, artificial intelligence, context free characteristics, questionnaire

# Kivonat

A kérdőíves adatgyűjtés esetén fennáll annak kockázata, hogy az előállt adatvagyon minősége kritikus, azaz nem különbözik érdemben a véletlentől a válaszadói intuíciót terhelő problémák miatt. Azonban minél inkább mérést helyettesítőek a kérdőív-kérdések, annál nagyobb az esélye a hiteles adatvagyon keletkeztetésének. Ahhoz, hogy nem mérés jellegű kérdésekre kapott kérdőív-válaszok értékét értelmezni lehessen, mérés-jellegű kérőív-kérdések válaszai (X-tömb: sportolás gyakorisága, alvászavar/dohányzás léte, stb.) alapján kell modellezni a nem-mérés-jellegű kérdőív-kérdések kapott válaszokat. Emellett a modellezés minden lépését végre kell hajtani véletlenszerű X-tömb alapján is. Egy MI-orientált és egyben automatizálható modellezésre alapozó adatvagyon-minősítés keretében elvárható, hogy a véletlenszerű adatvagyon minőségi mutatói rosszabb összképet mutassanak, mint a valós (kérdőíves) adatvagyon mutatószámai. Amennyiben a nyers adatvagyon idősoros jellegű, úgy a klasszikus regresszió-alapú és/vagy korreláció-alapú adatvagyon-értékelés nem kellően szofisztikált. Önálló cikkeket igényel majd a jövőben az ideális adatvagyon fogalmának pontosítása, ill. a korreláció dinamikus értelmezéséből fakadó problémák feltárása –a klasszikus adatvagyonminősítés további kockázataiként. Az adatvagyon-minősítés elsődlegesen tartalom-független illik, hogy legyen, de a tartalomfüggő aspektusok bemutatása is szerves része a konkrét esettanulmánynak. A vizsgált adatvagyon kapcsán nem-mérés-jellegű X-tömb is képezhető az alvásminőség különféle zavaró tényezőinek szubjektív megítélése mentén. Önálló cikként fog bemutatásra kerülni a kétféle X-tömb hatása a modellezés jóságára. Eredmény: a vizsgált adatvagyon fókuszált rétege jobb minőségű, mint a véletlen számok által képzett adatvagyon.

# Abstract

In the case of questionnaire data collection, there is a risk that the quality of the data generated is critical, i.e. it does not differ significantly from chance due to problems affecting the respondent's intuition. However, the more the questionnaire questions are a substitute for measurement, the greater the chance of generating authentic data. In order to interpret the value of questionnaire responses to non-measurement questions, the responses to non-measurement questionnaire questions must be modelled based on the responses to measurement-type questionnaire questions (X-array: frequency of sports, sleep disorders/existence of smoking, etc.). In addition, all steps of the modelling must also be performed based on a random X-array. In the framework of an AI-oriented and at the same time automatable modelling-based data asset evaluation, it can be expected that the quality indicators of random data assets will show a worse overall picture than the indicators of real (questionnaire-based) data assets. If the raw data assets are time-series in nature, then the classic regression-based and/or correlation-based data asset assessment is not sufficiently sophisticated. In the future, it will be necessary to clarify the concept of ideal data assets and to explore the problems arising from the dynamic interpretation of correlation – as additional risks of classic data asset evaluation. Data asset evaluation should be primarily content-independent (context free), but the presentation of content-dependent aspects is also an integral part of the specific case study. In connection with the examined data assets, a non-measurement-type X-array can also be formed along the subjective assessment of various disturbing factors of sleep quality. The impact of these two existing types of X-arrays on the quality of the modelling will be presented as a separate article. Result: the focused layer of the examined data set is of better quality than the data set generated by random numbers.

# Bevezetés

Ebben a cikkben Csizmár Attila (2025a) diplomamunkája (vö. *Az obstruktív alvási apnoe (OSA) gyakoriságának és komorbiditásának vizsgálata a Magyar Honvédség személyi állománya körében 2011-2022 között – Semmelweis Egyetem*) és IKEB-jelentése (Csizmár:2025b) keretében feldolgozott önbevallásból származó egészségügyi adatok alapján vizsgálatra kerül az általános érvényű adatvagyon-minőségbiztosítási hipotézis: Lehet-e minden év által érintett állománytól kapott adatvagyon másként egyformán hiteles?

Maga a Csizmár-féle diplomamunka (vö. 24. oldal) tételesen utal arra, hogy: „*Az önminősített egészség a globális, általános egészségi állapot szubjektív indikátora, mely a kutatási eredmények alapján jó prediktív értékkel bír az egészségi állapot objektív mutatóival kapcsolatban*” (vö. Hornyák, B. és Ördögh, I. (2015). Obstruktív alvási apnoe szindróma (OSAS) a Magyar Honvédség személyi állománya körében II. Az OSAS gyakorisága és komorbiditása a Magyar Honvédség személyi állományának körében. Hadtudományi Szemle, 8(3), 101-110.). Az önminősített egészségi állapot jellemzésekor a vizsgálati személyeknek a „Hogyan ítéli meg egészségi állapotát általában?” kérdésre egy 5 fokozatú skálán kellett válaszolniuk. A kérdésre adható válaszok: 1=Nagyon rossz, 2=Rossz, 3=Kielégítő, 4=Jó és 5=Kiváló voltak. Értelemszerűen a mindenkori válaszadók összlétszámát 100%-nak tekintve az 5 válaszréteg százalékos megoszlásának összege kényszerűen minden évben 100%-ot kellett, hogy eredményezzen.

Hasonló állítások fogalmazódnak meg Hornyák doktori értekezésében (Hornyák, 2022) és diplomamunkájában (Hornyák, 2011): vö. „*Az önminősített egészség a globális, általános egészségi állapot szubjektív indikátora, mely a kutatási eredmények alapján jó prediktív értékkel bír az egészségi állapot objektív mutatóival kapcsolatban (**Ilder, Benyamani, 1997; Burstroem, Fredlund, 2001), így az egészségi állapot és a MÁQ közötti kapcsolat vizsgálatánál az objektív (úm. betegség előfordulása) mutatók mellett a szubjektív mutatók alakulását is vizsgáltuk. Kutatási eredményeink szerint a magas MÁQ értékkel rendelkezők szignifikánsan jobbnak ítélték meg egészségi állapotukat (Hornyák, 2011).*” (MÁQ = Mentális Állóképesség Teszt).

A kérdőívezés megkerülhetetlenségét erősíti pl. az alábbi mű is (21 említéssel a kérdőív-kulcsszó kapcsán): Szilágyi, Zsuzsanna; Csukonyi, Csilla; Sótér, Andrea; Hornyák, Beatrix: A mentális állóképesség-vizsgálatok bevezetésének lépései a Magyar Honvédség állományában (2006-2009) HADTUDOMÁNYI SZEMLE 7 : 1 pp. 158-178. , 21 p. (2014) - <https://tudasportal.uni-nke.hu/xmlui/static/pdfjs/web/viewer.html?file=https://tudasportal.uni-nke.hu/xmlui/bitstream/handle/20.500.12944/10608/2014_1_tt_szilagyi.pdf?sequence=2&isAllowed=y>

A szerzők kapcsolati hálójában is tetten érhető az idősoros kérdőív-adatok elemzésének gyakorlata, kultúrája: vö. Novák, Attila; Hornyák, Beatrix; Rázsó, Zsófia; Szalánczi, Szabolcs; Sótér, Andrea; Juhász, Zsolt; Nyakas, Csaba (2019) - The introduction of health behavior profiles in the Hungarian Defense Forces: a cluster analysis of lifestyle factors according to the health screening tests performed in 2011–2015 - INTERNATIONAL JOURNAL OF OCCUPATIONAL MEDICINE AND ENVIRONMENTAL HEALTH 32 : 1 pp. 99-114., 16 p. - <https://pubmed.ncbi.nlm.nih.gov/30855101/>

Jelen cikk hipotézisvizsgálata tehát lényegében egy robot-kutató azon feladatát demonstrálja, aki, mielőtt elemzi egy adott adatvagyon adatait, objektív (módszeres) lépéssorral meg kíván győződni arról, hogy az általa elemzendő adatvagyon is megfelel-e a Csizmár-Hornyák-Ördögh-Ilder-Benyamani-Burstroem-Fredlund tapasztalatoknak, melyek bár általános érvénnyel bírónak látszanak, de magától értetődően bármely konkrét adatvagyon esetén előállhat, hogy a mindenkori válaszadók válaszai véletlenszám-generálás eredményei(hez hasonló minőségűek).

Az adatvagyonok értékének becslése a cikk fókuszát jelentő mesterséges intelligencia-alapú lépéssor mellett Copilot-bevonására támaszkodva is megtörténik. Értékbecslésről abszolút értelemben nem szokás beszélni, de egy benchmarking folyamat keretében az értékesség fogalma operacionalizáltan értelmezhetővé válik. A benchmarking folyamat minimális eleme kell, hogy legyen egy véletlenszerű adatvagyon létrehozása. (Külön cikk foglalkozik a jövőben az értékskála másik végpontjával, vagyis az ideális adatvagyon kialakításának módszertanával.)

A cikk által bemutatott értékelési módszertanok kényszerűen context-free jellegűek kell, hogy legyenek, vagyis az értékességről a vizsgált jelenségektől alapvetően független lépéssor alapján kell tudni nyilatkozni.

# Szakirodalmi előzmények

Bár a vizsgálandó adatvagyon emberi egyedek egészségügyi/életvezetési adatait vizsgálja, de szakirodalmi előzményként a tartalomfüggetlenségi elvárás okán csak és kizárólag módszertani előzményekre kell és érdemes utalni:

* Kérdőív-kérdések kapcsolatának feltárása (vö. Pitlik László: Hogyan leplezhető le modellezéssel/ChatGPT-vel, hogy egy kérdőív mely két kérdése mond leginkább ellent egymásnak? – 2025: <https://miau.my-x.hu/miau/319/kerdoiv.xlsx>)
* Adatvagyon-érték COCO STEP-IX (Component-based Object Comparison for Objectivity) keretében történő letapogatása: (vö. Angyal János: Szerverek MI-támogatással optimalizált ár/teljesítmény-alapú beszerzése – 2024: <https://miau.my-x.hu/miau/315/aj_plus.pdf>)
* Fizikai jelenségek közötti kapcsolatok quasi hibátlan modellezhetőségének feltárása (vö. Pitlik Marcell: Térfogati égés optikai vizsgálata és mesterséges intelligencia alapú elemzése – 2021: <https://miau.my-x.hu/miau/277/szakdoga_PM_final.pdf>), ill. Pitlik Marcell: Égések zaj alapú elemzése mesterséges intelligenciával – 2021: <https://miau.my-x.hu/miau/279/tdk_zaj/TDK_PM_zaj_final.pdf>)
* Komplex jelenségek közötti kapcsolatok feltárása modellezés keretében (vö. Németh István, Szani Ferenc, Pitlik László: Tananyagok összehasonlító elemzése szemmozgás/érzelem/energia/EEG és egér/billentyűzet-log-adatok alapján - 2022: <https://miau.my-x.hu/miau/288/2d_3d_complex_study.docx>)
* Kérdőíves adatvagyonok hazugságpotenciáljának feltárása (vö. Pitlik László, Pitlik Marcell: Hazugságvizsgálat kérdőívezés keretében - 2019: <https://miau.my-x.hu/miau/256/torrent/liar-detection-in-questionnaires.docx>)
* Inkonzisztencia-alakzatok feltárása kérdőívezek segítségével (vö. Pitlik László: Inkonzisztencia-vizsgálatok lehetősége kérdőíves válaszok alapján – 2017: <http://miau.my-x.hu/miau/232/elegedettsegi_hermeneutikak_v2.docx>)
* Gyanúgenerálás kérdőívek alapján: (vö. Pitlik László: Nominalskaliert interpretierte Befragungen in der Verdachtsgenerierung – 2012: <http://miau.my-x.hu/miau/162/93_3.docx>) + <https://miau.my-x.hu/miau/159/93_2.docx>)
* Hazugságvizsgálat kérdőíves alapokon: (vö. Pitlik László Horváth Csaba, Szabó László: Pénzügyi tudatosság más szemmel, avagy hazudj, ha tudsz a kérdőívezés során – 2011: <http://miau.my-x.hu/miau/158/la158.docx>)

A hivatkozott előzmények saját hivatkozásai is szerves részét képezik a kérdőívek, mint objektum-attribútum-mátrixok kapcsán kialakított általános érvényű ellenőrző modellezési logikának, ahol minden egyes kérdőívválaszra becslés készítendő a többi adat alapján, majd a minden tény-becslés-eltérés kérdésenkénti és/vagy kitöltőnkénti anti-diszkriminatív értelmezése vezet el a leginkább kritikus kérdések és/vagy válaszadók feltárásához.

A context free elemzés mellett természetesen a tartalomfüggő kérdéskapcsolatok problémája is létező és hasonló logikák mentén kezelhető.

# Copilot tapasztalatok

A szakirodalmi előzmények speciális kiegészítő rétege manapság az, ha a vizsgált problémát megkíséreljük megoldatni LLM szolgáltatásokkal (pl. Copilot-tal).



1. Ábra: Copilot-alapú adatvagyon-minősítés (forrás: saját ábrázolás)

Az 1. ábra (egyelőre az szcenáriók keletkezéstörténetének ismertetése nélkül, de) világos útmutatást ad arra vonatkozóan, hogyan lehet a Copilot szerint az X-tömb és az Y tömb közötti összefüggés-idealitásról algoritmikusan nyilatkozni. (A pirossal jelzett számok/karakterek arra a furcsa anomáliára utalnak, miszerint a Copilot-vélemény mögött valódi számítások történnek, melyek még sem konzisztensek, vagyis már az átlagszámítás is lehet a kerekítésen túl is aggályos esetlegesen.



1. Ábra: Copilot válasz azonos prompt esetén (forrás saját ábrázolás)

Az 1. és a 2. ábra összevetése rámutat arra, hogy a Copilot saját magához képest mérve sem következetes: az 1. és 4. szcenárió azonos R2 értékre vezet, a 2. szcenárió R2 értéke a részeredmények alapján számított értékkel azonos, míg a 3. szcenárió esetén a három részeredmény jelentősen eltér egymástól (0.9472 vs. \*77 vs. \*78).

Ezek a számszakilag létező (de mennyiségi szempontból apró) hibák a lényegen nem változtatnak: az 1. és a 4. szcenárió közül az 1. (a kérdőíves valóság) a jobb, míg a 2. (a véletlenszám-alapú) benchmark a rosszabb. A 3. és a 4. szcenárió az ideális X-tömb külön cikkben bemutatandó részeredményei, melyek nem jobbak a Copilot szerint, mint a valóságban realizált kérdőíves adatok.

Az R2 mellett az RMSE értékek alapján is egyértelmű és az R2-hez mérten azonos a sorrend.

Azonban a korrelációk alapján (vö. 3. ábra):



1. ábra: Y-rétegek korrelációi az X-tömb változói esetében (forrás: saját ábrázolás – szcenáriók fentről lefelé: #1-2-3-4)

a győztes (ideál?) szcenárió már a 4. szcenárió, ill. az 1. (valós) szcenáriós jobb, mint a 2. (random) szcenárió, hiszen az átlagos előjelmentesített korrelációs értékek maximuma: 0.83, minimuma 0.36.

# Adatvagyon

A feldolgozott adatvagyon idősoros: a 2011-2022 közötti éveket (11 db) érintik a kérdések. A 2014-es adatok technikai okok miatt nem kerülhettek egyelőre feldolgozásra. A 2018-as év rekordszáma töredéke (1:10) a többi év rekordszámának – ennek oka is további tisztázásokra vár. Ellenben ez a mennyiségi eltérés mindenképpen inicializálta a százalékos adatértelmezést, vagyis azt, hogy a vizsgált 11 attribútum esetén évről évre a mindenkor választ adók összes számát kelljen 100%-nak tekinteni (vö. 1. ábra).

A 11 attribútum 4 kérdéshez, s kérdésenként 2-3-4 opcióhoz kapcsolódik:



1. ábra: X-tömb (forrás saját ábrázolás)



1. ábra: Y-tömb (forrás: saját ábrázolás)

Ahogy a 4. ábrán látható, a fizikai aktivitás, mint input bináris, hasonlóképpen az alvászavart érintő jelenséghez. A dohányzás 3 opciós, a sport 4 opciós. Minden X-jelenség quasi mérhető, a klasszikus önbevallás teljesen szubjektív kihívásaitól alapvetően mentes (vö. Mennyire szereti Ön a pirézeket? 1<5). A quasi mérhető, de kérdőíves önbevalláson keresztül keletkező adatokkal szemben az 5. ábra, vagyis az Y-tömb tipikusan a szubjektivitás „tombolásaként” értelmezhető. Ezért is racionális a Csizmár-Hornyák-Ördögh-Ilder-Benyamani-Burstroem-Fredlund-vélelem vizsgálata ezen konkrét adatvagyonra vonatkozóan.

A teljes adatvagyon a 11 év alatt 20.000+ személyt érintett, így, ha reprezentativitásról nem is lehet beszélni jelen vizsgálatok kapcsán, de egy masszív tehetetlenségi nyomatékról, az éves arányok attribútumonkénti arányok stabilitásáról igen.

A korrelációs értékek (vö. 3. ábra) ismét csak azt erősítik, hogy az X-tömb és az Y-tömb között masszív összefüggések vannak. De a korreláció-alapú vizsgálat már elveszíti idősoros jellegét, vagyis korrelációs alapon nem lehetne arról beszélni, vajon lehet-e minden év másként egyforma? Más szavakkal: az egyre változó személyállomány őszintesége, intuíciós potenciálja változik-e a vizsgálható 11-12 év során?



1. ábra: A regressziós Copilot számítások részletei (forrás: saját ábrázolás)

A 6. ábra felfedi a regressziós számítások struktúráját, vagyis minden egyes Y-réteg 11 X(i) alapján 11 (éves) ismétlésben áll elő, vagyis a regressziós számítások esetén is elvész az évek hasonlóságának összehasonlíthatósága!

# Módszertan

Most, hogy ismertek az kérdőíves adatvagyon releváns részletei és ismertek a szakirodalminak (mesterséges intelligencia mentesen rendelkezésre álló) módszertani alternatívák, máris belátható volt, hogy az évek összehasonlítására alkalmatlanok a klasszikus értelmezések. Már pedig a kérdőívek adatvagyonának értékelésekor nem elegendő csak a teljes adatvagyont értelmezni, ha ennek idősoros rétegei mentén komoly kockázatok állhatnak elő. A hipotézis tehát egyszerű: lehet-e minden év másként egyformán modellezhető? Ez a hipotézis elsődlegesen egy MI-alapú, anti-diszkriminatív optimalizálást igényelne (vö. COCO-Y0: <https://miau.my-x.hu/miau/196/My-X%20Team_A5%20fuzet_HU_jav.pdf>), de mint látni fogjuk, abban az esetben, amikor a véletlenszámok alapján történő modellezés is „meglepően” jó részeredményekre vezet, s következésképpen a valós adatvagyon kapcsán szinte keresni kell a modellezési hibákat, felesleges ágyúval verébre lőni. A folytatást jelentő cikkek kapcsán, ahol az ideális adatvagyon mibenlétének feltárása érdekében további szcenáriók kerülnek kialakításra, az anti-diszkrimintív modellezés megkerülhetetlenné válik, hasonlóan a regressziós együtthatókból következő korreláció-fogalom értelmezésének kockázataihoz, ill. a nem-mérés-jellegű X-tömb értelmezéséhez.

A regresszió 20 modellje a Copilot alapján (vö. 6. ábra) a 4 szcenárió 5 Y-rétege kapcsán (4\*5) adatkonstelláció értelmezhető. Ezekben az OAM-okban (objektum-attribútum-mátrixokban) a 11 év nem értelmeződik idősorként. Csak a 20 OAM-nak vannak mutatószámai a regressziós lépések elvégzésének eredményeként. A helyzet hasonló a korrelációs értékek esetén is (vö. 3. ábra), ahol lényegében még az Y-rétegek sem kell, hogy elválasztásra kerüljenek egymástól. A regressziós számítások kapcsán is a tény vs. becslés korrelációk (R2-értékek) Y-rétegenkénti átlaga vezet a szcenáriók versenyének értékeléséhez. Ez nem mellesleg felveti azt a kérdést is: miért is volt szükség a regressziós számításokra, különösen a linearitás feltételezésére? Más megfogalmazásban: a linearitás önkényessége helyett lett-e volna értelme más függvényalakot választani? S még általánosabb szintre emelve a kritikát: Létezik-e az emberiség tudásában az a modul, mely képes a regresszió lehetőségei közül a legjobbat meghatározni? (A válasz erre a meta-szintű kérdésre ismét csak önálló cikket igényel…)

A lehet-e-minden-év-másként-egyformán-modellezhető-hipotézis kapcsán olyan modellezési keretrendszerre van szükség, mely rugalmasabb a regressziós számításoklogikájánál. A lépcsős függvények ilyen alakzatok, hiszen a regresszió „Y=a\*X+b” alakját jellemző „a” paraméter helyett a bemeneti jelek rangsorszintjeihez egyenként kerül egy paraméter meghatározásra, vagyis az inputadatoknak nem a valós nagyságrendje, csak sorszáma kerül feldolgozásra vezérlő jelként. (Az, hogy itt történik-e tényleges információvesztés, ismét csak egy önálló cikket igényel…)

Miért szükséges a rugalmasabb modellezési keretrendszer? Egyrészt, mert a regresszió kapcsán az egyes inputváltozók „a” értéke előjeles számként áll elő, vagyis a regressziós modellezés maga szeretné eldönteni, milyen hatásmechanizmust vélelmezzen Xi és Y között. A lépcsős függvények esetén erről a modellező dönt. S éppen ez a döntés vezet el az ígért további cikkek egyikéhez, ahol is a Y vs. Xi korreláció, vagyis az egyenes vagy fordított arányosságok vezérlőjelkénti kiaknázásában rejlőcsapdák kerülnek felmutatásra.

Másrészt minél több paraméterrel dolgozik egy modellezési keretrendszer (vö. mesterséges neurális hálók), annál nagyobb az esélye annak, hogy a becslések és a tények azonossá válnak. Erre pedig azért van szükség, mert a véletlenszámokból álló nyersadatok esetén egy eleve rugalmatlan modellezési keretrendszer (vö. regresszió, különösen ennek lineáris alakzata) szinte minden megfigyelés (év) esetén hibás becsléseket fog adni vélhetően még az egyszer később feltárandó ideális adatvagyon esetén is, vagyis fel sem tud merülni az az eredmény, hogy a valós és a véletlenszámokra alapozott OAM-szcenáriók mindegyike hibátlan tény vs. becslés állapotokra vezessen. A lépcsős függvények rugalmassága esetén ennek esélye, amint hamarosan a részeredmények alapján tételesen látható is lesz: magas.

Ebben a cikkben tehát annak a párhuzamos vizsgálatnak az eredményei kerülnek bemutatásra, ahol az 1. ábra 1. és 2. OAM-szcenáriói kerülnek az 5 Y opciónak megfelelően 5-5 lépcsős függvénnyel modellezésre úgy, hogy a nyersadatok kapcsán a szakértői szintű sorszámozási paraméterek (irányvektor) a következő egyenes, ill. fordított arányosságokat kényszerítik ki a regressziós számítások szabad előjel-paraméterei helyett:



1. ábra: A szakértői szinten meghatározott irányvektor (forrás: saját ábrázolás)

A 7. ábra a figyelmes szemlélő számára egyetlen két kritikus ponttal rendelkezik: sportolás-heti-2-3-alkalommal (x4), ahol a szakértői iránykód = 0, vagyis minél nagyobb a válaszadók ezen kategóriába eső aránya (%) az összes válaszadó függvényében, annál nagyobb illene, hogy legyen bármely Y-rétegbeli arány. A korrelációs értékek az Y5, vagyis a kiválóság arányát jelentő Y-rétegre vonatkoznak (ellenőrzésképpen: vö. 3. ábra legfelső kép-réteg y5 oszlopai – x3-4-5-6 változók esetén). Ezen kritikus változó esetén igaz, hogy az egyenesarányosság szakértői elvárásához (a 0-s iránykódhoz) negatív (masszívan negatív = -0.94) korreláció tartozik. A másik kritikus változó az x6-változó, vagyis a skála egyik széle.

Miért kellett ignorálni a korrelációs számítás x4 esetén? A válasz egyszerű és részlegesen kontextus-függő: ha egy négyelemű skála nem szélső pozícióban lévő elemének viselkedése eltér a szomszédos elemek viselkedésétől, akkor az logikátlannak minősítendő (vö. polinomok vs. lineáris regresszió értelmezhetősége). Amennyiben tehát x4 nem esik az opciólista (1-2-3-4) valamely szélére, akkor nincs ok a másik három elem karakterisztikájától eltérni!

A fizikai aktivitás binárisan ábrázolt ténye és/vagy az alvászavar hasonlóan bináris alakzata esetén az egyik opció a másik ellentéte, s ezt erősíti a korrelációs érték is mindkét esetben – kényszerűen, lévén a bináris alakzat egyik értékének ismeretében a 100-egyik=másik elv kényszerpályára terel minden további matematikai/logikai műveletet.

Miért nem érvényes ezen kényszerpálya a sportolási gyakoriság négy-elemű opciólistáján belül x6-ra? A válasz nem egyszerű: először is, lehetne érvényes pl. az 1-2-3 opció lehetne egyenes arányosságot tükröző, míg a 100-egyik-másik-harmadik=negyedik elv mentén a 4-es opció (x6) lehetne fordított arányosság. DE: a sportolási gyakoriság 4 fokozata közül a legkevesebb sem minősül a jelenség (sportolás) egyértelmű tagadásának, mint a bináris esetekben. Így, ha x4 korrelációja kényszerűen felül bírálandó, akkor x6 esetében a három további opció immár homogén iránykódja öröklődik a negyedik opcióra is.

Miért kell a korrelációkat elsődlegesen y5 esetén vizsgálni, noha a 3. ábra felső rétegében x3-4-5-6 esetén adott a többi korrelációs érték is y4-3-2-1-re is? A válasz relatíve egyszerű: ha az y5 esetét tekintjük úgy, hogy az egy szélsőségesen jó (kiváló) érték, akkor minden más ennek ellentettje, tehát a további rétegek eredője számítana csak, ami a 100-egyik=másik elv alapján kényszerpálya. Tehát elegendő a kiváló opciót (y5) alapul venni.

S végül nincs-e ellentmondás a dohányzás három-opciójú esetében a fenti okfejtések egyidejű érvényesítése mellett? A válasz egyszerű: nincs. Miért? Mert a dohányzás három opciója közül a soha nem dohányzás ellentettje a másik kettő, melyek egymás fokozatai. Vagyis az ellentettség okán, ha a soha opció iránya korrelációs megerősítés mellett 0-ás, akkor a másik két (ellentett) pozíció iránykódja kényszerűen 1-es, amit a korrelációk is megerősítenek.

A részleges kontextus-mentességre utalást indokolja a fentiek alapján, hogy egyetlen egy logikai művelet sem használta a sport, a dohányzás, stb. kulcsszavakat, csak az opciólistaelemek számosságát és ezek ellentett jellegét, ami végtelen sok további jelenségre igaz lehet, de a jelenségek valamilyen specialitása mégis csak értelmezésre került…

A lépcsős függvények paramétereinek levezetése a MY-X FREE COCO-online szolgálatással történt: <https://miau.my-x.hu/myx-free/>, ill. <https://miau.my-x.hu/myx-free/coco/index.html>.

# Eredmények

A valós adatok alapján a rugalmas lépcsős függvények mind az öt Y-réteg esetén minden évre vonatkozóan nulla hibás becslést engedtek meg – a (magas korrelációk alapján és a jó adatvagyonra vonatkozó) elvárásoknak megfelelően (vö. <https://miau.my-x.hu/miau/326/s1c1/s1c1_overview.xlsx>).

Ezzel szemben a véletlenszerű adatvagyon kapcsán modellezési hibákat kellene kimutatni tudni a szakértői szinten (vö. a módszertani fejezetben) rögzített iránykódok szisztematikus betartása mellett. DE az X-tömb esetén (az jelenségek 2-3-4-elemű opcióhalmazainak szabályszerűségeit betartva) generált véletlen számok alapján is igaznak bizonyult, hogy mind az öt modellréteg hibátlanul becsülhető.

Vagyis a Copilot regressziós (R2-alapú és/vagy modellezési hiba-orientált), ill. a korrelációs alapú megközelítései alapján vélelmezett szcenárió validitási, konzisztencia-sorrend helyett a rugalmasabb modellezési keretek arra hívják fel a figyelmet, hogy a kérdőíves adatvagyon a szakértői vélemény rögzítettsége esetén nem rendelkezik triviális értéktöbblettel a véletlenszerű adatvagyonhoz képest. Hogy a szakirodalmi helyzetértékelés gyanúsan (naivan) egyértelmű, arra a korreláció-alapú nézet hívta fel a figyelmet, ahol is az ideálnak remélt egyik (3. vagy 4.) szcenárió mindegyike legyőzte a valós (kérdőíves: #1) és különösen a véletlenszám-alapú (#2) szcenáriót (vö. 3. ábra), míg a regresszió-alapúság esetén a #3 és a #4 középre szorult a valós és a véletlen szcenáriók közé.

Azonban:

Továbbra is fenntartva, hogy a regresszió kapcsán fókuszált paraméter-előjelek kérdésköre önálló cikket kíván, ezen a ponton a valós kérdőíves adatvagyon értékének mibenlétét érintő bizonyítási folyamat további lépésekkel egészítendő ki (vö. Melléklet: Copilot önvezérlő hibajavító logikája jelen példa kapcsán megtapasztalva):

A következő lépés értelemszerűen a dinamikus iránykódokkal folytatott modellezés lehet és kell, hogy legyen, vagyis ahol az iránykód mindenkor megfelel a mindenkori Y-réteg alapján a fix X-tömbhöz változónként számított korreláció előjelének. Minden más módszertani lépés azonos maradt.

Eredményként az alábbi tény vs. becslés állapotok álltak elő:

A valós (kérdőíves) modellsorozat esetén az y5-4-3-2 rétegekben a modellezés továbbra is hibátlan, de az y1 esetén megjelenik az X-tömbben az első antagonizmus, vagyis egy olyan helyzet, ahol már az input-adatpárok kapcsán az egyik év (objektum) adatpárjai minden változó esetén jobbak, ill. soha nem rosszabbak egy másik évvel összevetve (vö. 2011 vs. 2021) és ezzel párhuzamosan az Y oldal párösszehasonlítása nem követi az X-tömbbeli előnyösség szisztematikus alakulását. Az antagonizmus tipikus alapja a sorszámozott OAM-ok esetén a sorszámokra vonatkozik. Az atipikus alakzatok közé a sorszámok értékeire vonatkozó párösszehasonlítások tartoznak: jelen esetben az x5 esetén (vö. 2011: 9. hely, 2021. 7. hely – minden más esetben a 2011-es y1 nagyobb illene, hogy legyen, mint a 2021-es y1, míg a valóságban 2011-es y1 érték < 2021-es y1 érték).

A véletlenszámokra alapozó X-tömb esetén az antagonizmusok már három Y-rétegben is megjelennek (y1-2-3), míg az y4-5 továbbra is hibátlan tény vs. becslés állapotpárokat ad minden évre. Az y2 és az y3 rétegben egy-egy tipikus és atipikus antagonizmus érhető tetten, míg az y1 esetén a becslés vs. tény korreláció maga zuhan már be 0.54-es szintre (vö. y2-3: korreláció = 0.99, y4-5: korreláció = 1.00).

# Vita

A szakirodalmi (Copilot-javaslatot követő) regressziós, ill. korreláció-alapú naiv, mindennemű objektivitást nélkülöző számítások látszólag triviális és megnyugtató eredményre vezettek a valós (kérdőíves) és a véletlenszerű nyersadat-szcenáriók jóságversenyét illetően. De ezen módszerek mentén sem az évek nem hasonlíthatók össze, sem az ideális X-tömbre vonatkozóan nem lehet állástfoglalni, hiszen a próbaként MS Excel Solver-rel levezetett szcenáriók (#3: csak y5-re vonatkozóan legyen minden x1-x11 korreláció 1.00, ill. #4: minden Y-réteg esetén a korrelációk abszolút értékeinek átlaga legyen maximális) értékelése a regresszió-alapúság és a korreláció-alapúság esetén karakteresen eltért egymástól. Vagyis nincs ideál-definíció a szakirodalom alapján automatikusan adva.

Az iránykódok dinamikus vezérlőjelként való felhasználása nem idegen a regressziós közelítéstől, sőt, lényegében a regresszió fogalma alapozza meg az iránykód korrelációfüggőségét. Éppen ezért a lehet-e minden év másként egyformán racionálisan modellezhető probléma érzékenység-vizsgálatának tekinthető az az eljárás, amikor a rögzített szakértői iránykódok esetét dinamikus iránykód-vezérlésre cserélük le.

Az ideális X-tömb kihívásához hasonlóan létezik az adatvagyon-minősítő skála másik végét érintő másik kihívás is: mit is jelent a legvéletlenebb véletlen fogalma? Hiszen véletlenül a valós (kérdőív-alapú) X-tömb, sőt a későbbiekben feltárandó ideál-verzió(k) is legenerálhatók. A véletlenszerűség mibenléte kapcsán érdemes már itt utalni arra, hogy pl. a világűr entrópiája nem akkor a legszélsőségesebb, ha benne az anyag/energia quasi homogén módon oszlik el – mert a homogenitás is egy fajta mintázat.

Úgy az ideális X-tömb, mint a véletlen X-tömb kapcsán felmerülhetne teoretikusan a brute force alapú (kombinatorikai jellegű) alternatíva-generálás lehetősége. Ebben a pillanatban quasi végtelen sok alternatív X-tömböt kellene értékelni modellezési lépések eredményei alapján. Az anti-diszkriminatív modellezés képes arra az ideális és a véletlenszerű minták, vagyis quasi az összes(nek vélt) X-tömbvariáns esetén, hogy egyetlen egy skálára konvertálja az X-tömbök idealitását (függvény-szimmetria-alapon önellenőrzött módon), ahol a preferált modell-részeredmények jelentik az egyik (ideális) végletet, ezek tükörképe a másik (véletlen) végletet. A többi minta a normaszerűség megtestesítése. A valós (kérdőíves) adatvagyonok kapcsán elvárható, hogy norma felettiek (azaz ideális jellegűek legyenek). Az önellenőrzés keretében a mindenkori OAM-méret függvényében beszélhetünk egy negyedik értelmezési típusról is: ezek a nem validálható X-tömb-mintázatok, melyek a szélsőségesen ideálistól, a szélsőségen véletlennek tűnőig előfordulhatnak bárhol bármilyen arányban, s azt jelzik, hogy a kellően rugalmas modellezési keretek mellett ezek értékéről adott adottmennyiség és -tartalom esetén még kockázatos nyilatkozni.

A modellezés keretrendszere az itt és most alkalmazott COCO-STD-hez (termelési függvény-generáláshoz) képest is növelhető a COCO MCM (exploratív) eljárással, mely akkor szükséges, ha az iránykódok meghatározását nem lehet logikai úton elvégezni. Ahogy a 0.54-es korrelációra vezető dinamikus korrelációs vezérlőjellel dolgozó y1-es modell is jelzi, nagyobb OAM-ok és/vagy komplexebb x-x kapcsolatok esetén a modellezés minden Y-rétegre alacsony korrelációs becslési hiba-mintázatokat fog adni, melyek egymással szembeni idealitását/előnyösségét ismét csak a COCO Y0 (anti-diszkriminatív) modellezés fogja skáláva transzformálni tudni.

# Konklúziók

A szakirodalmi megoldások naivitása, vagyis egy anti-diszkriminatív ellenhipotézis (lehet-e minden év másként egyforma?) megfogalmazásának képtelensége önmagában is kritikussá teszi a szakirodalom által sugallt módszertani önkényt. Az önkényesség (naivitás) további lelepleződését jelent az ideál hiánya, ill. a potenciális X-tömre vonatkozó ideálok ellentmondásos értékelései. Arra a kérdésre tehát, hogy lehetünk-e elégedettek a szakirodalmi alapokkal (vö. a Copilot-válaszokkal), a komplexitás magasabb szintjén a válasz: NEM.

Az a tény, hogy véletlenszámok alapján is lehetséges a nem mérés-orientált kérdőívválaszok (y1-2-3-4-5) modellezés keretében történő hibátlan becslése, kijózanító kell, hogy legyen minden klasszikus statisztikai értelmezésben járatos (s főleg az abban még csak nem is járatos) elemzők, olvasók számára. Vagyis a közgazdaságtan kapcsán szokásos érzékenységvizsgálatok logikáját a modellezés, hipotézisértelmezés világára is érdemes kiterjeszteni: jelen esetben az azonos méretű, de véletlenszerű adatvagyonokra vonatkozó azonos módszertannal előálló eredmények ismerete nélkül nem illik a valós adatvagyonokról nyilatkozni. Ennek az elvárásnak egy speciális szelete az ideál fogalmának KNUTH-i szintre konvertálása (vö. [https://miau.my-x.hu/miau2009/index\_tki.php3?\_filterText0=\*knuth](https://miau.my-x.hu/miau2009/index_tki.php3?_filterText0=*knuth)).

A látszólag a véletlennel egyenértékű valós adatvagyon potenciális értékességét (egyelőre még az ideális adatvagyon fogalmi és a korrelációvezérelt modellezés kockázatai nélkül, de mindenképpen) a korreláció-vezérelt modellezés keretei között az antagonizmusok fellépése érzékletesen mutatja be. A véletlenszerű adatvagyonok megfelelő szögből szemlélve klasszikus/tipikus antagonizmusokat mutatnak fel, amit egy valós/értékes adatvagyon nem illik, hogy felmutasson. Természetesen az atipikus antagonizmus is kockázati tényező, de ennek fellépését magyarázó gondolatok már az ideál-definíciót kifejteni akaró cikkbe tartoznak. Itt és most csak annyit kell deklarálni, hogy olyan Y-tömb esetén, melyen belül belső törvényszerűségek adottak (vagyis az y1-2-3-4-5 rétegek összege mindenkor 100%), ott konstans X-tömb idealitásának léte, ill. az egyetlen ideális állapot helyett a rel. sok alternatív idealitás léte, mint hipotézis mindenképpen vizsgálandó, hogy nehogy ismét a naiv/önkényes értelmezések világába lépjünk vissza…

Az itt bemutatott termelési függvény-alapon automatizálható lépéssor a vita kapcsán említett validálási, anti-diszkriminatív, exploratív rétegekkel/modulokkal is kiegészíthető/kiegészítendő az univerzalitás növelése érdekében, hiszen a JÓSÁG fogalma végtelenül komplexé tehető/teendő…

# Hivatkozások

Csizmár, A (2025a). Az obstruktív alvási apnoe (OSA) gyakoriságának és komorbiditásának vizsgálata a Magyar Honvédség személyi állománya körében 2011-2022 között, Diplomamunka, Semmelweis Egyetem

Csizmár, A (2025b). Beszámoló kutatási eredményekről - Az obstruktív alvási apnoe (OSA) gyakoriságának és komorbiditásának vizsgálata a Magyar Honvédség személyi állománya körében 2011-2022 között, MHEK-IKEB-jelentés

Pitlik László (2025): Hogyan leplezhető le modellezéssel/ChatGPT-vel, hogy egy kérdőív mely két kérdése mond leginkább ellent egymásnak? – <https://miau.my-x.hu/miau/319/kerdoiv.xlsx>

Angyal János (2024): Szerverek MI-támogatással optimalizált ár/teljesítmény-alapú beszerzése - <https://miau.my-x.hu/miau/315/aj_plus.pdf>

Németh István, Szani Ferenc, Pitlik László (2022): Tananyagok összehasonlító elemzése szemmozgás/érzelem/energia/EEG és egér/billentyűzet-log-adatok alapján - <https://miau.my-x.hu/miau/288/2d_3d_complex_study.docx>

Hornyák, B. (2022): A mentális állóképesség mérésének és fejlesztésének lehetősége a Magyar Honvédség hivatásos és szerződéses személyi állományának körében, PhD-értekezés, Nemzeti Közszolgálati Egyetem, Hadtudományi Doktori Iskola, <https://hdi.uni-nke.hu/document/hdi-uni-nke-hu/Hornyak_Beatrix_doktori%20%C3%A9rtekez%C3%A9s.pdf>

Pitlik Marcell (2021): Térfogati égés optikai vizsgálata és mesterséges intelligencia alapú elemzése - <https://miau.my-x.hu/miau/277/szakdoga_PM_final.pdf>

Pitlik Marcell (2021): Égések zaj alapú elemzése mesterséges intelligenciával - <https://miau.my-x.hu/miau/279/tdk_zaj/TDK_PM_zaj_final.pdf>

Pitlik László, Pitlik Marcell (2019): Hazugságvizsgálat kérdőívezés keretében - <https://miau.my-x.hu/miau/256/torrent/liar-detection-in-questionnaires.docx>

Novák, Attila; Hornyák, Beatrix; Rázsó, Zsófia; Szalánczi, Szabolcs; Sótér, Andrea; Juhász, Zsolt; Nyakas, Csaba (2019) - The introduction of health behavior profiles in the Hungarian Defense Forces: a cluster analysis of lifestyle factors according to the health screening tests performed in 2011–2015 - INTERNATIONAL JOURNAL OF OCCUPATIONAL MEDICINE AND ENVIRONMENTAL HEALTH 32: 1 pp. 99-114., 16 p. - <https://pubmed.ncbi.nlm.nih.gov/30855101/>

Pitlik László (2017): Inkonzisztencia-vizsgálatok lehetősége kérdőíves válaszok alapján – <https://miau.my-x.hu/miau/232/elegedettsegi_hermeneutikak_v2.docx>

Hornyák, B. és Ördögh, I. (2015): Obstruktív alvási apnoe szindróma (OSAS) a Magyar Honvédség személyi állománya körében II. Az OSAS gyakorisága és komorbiditása a Magyar Honvédség személyi állományának körében. Hadtudományi Szemle, 8(3), 101-110., <https://epa.oszk.hu/02400/02463/00027/pdf/EPA02463_hadtudomanyi_szemle_2015_02_099-111.pdf>

Szilágyi, Zsuzsanna; Csukonyi, Csilla; Sótér, Andrea; Hornyák, Beatrix (2014): A mentális állóképesség-vizsgálatok bevezetésének lépései a Magyar Honvédség állományában (2006-2009) HADTUDOMÁNYI SZEMLE 7 : 1 pp. 158-178. , 21 p. - <https://tudasportal.uni-nke.hu/xmlui/static/pdfjs/web/viewer.html?file=https://tudasportal.uni-nke.hu/xmlui/bitstream/handle/20.500.12944/10608/2014_1_tt_szilagyi.pdf?sequence=2&isAllowed=y>

Pitlik László (2012): Nominalskaliert interpretierte Befragungen in der Verdachtsgenerierung I/II – <http://miau.my-x.hu/miau/162/93_3.docx> + <https://miau.my-x.hu/miau/159/93_2.docx>

Hornyák, B. (2011): Mentális állóképesség összefüggése az egészségmagatartással, Diplomamunka: DEOEC Népegészségügyi Kar Egészségpszichológiai Szak

Pitlik László Horváth Csaba, Szabó László (2011): Pénzügyi tudatosság más szemmel, avagy hazudj, ha tudsz a kérdőívezés során – <http://miau.my-x.hu/miau/158/la158.docx>

# Mellékletek

Részletek: <https://miau.my-x.hu/miau/326/s1c1>

Copilot önmaga által vezérelt hibajavítási folyamat bizonyítékai:

*„Úgy tűnik, hogy a nyers adatok feldolgozása során hiba lépett fel: az adatok oszlopainak száma nem egyezett meg a fejlécben megadott oszlopszámmal. Ez valószínűleg abból adódik, hogy az adatok tabulátorral tagoltak, és az első oszlop (pl. o1, o2, stb.) nem lett megfelelően kezelve.*

*A probléma megoldásához újra fogom strukturálni az adatfeldolgozást, hogy az objektumazonosítókat is figyelembe vegye, és minden oszlop pontosan illeszkedjen. Máris javítom és újra lefuttatom a fájl létrehozását.*

*A fájl létrehozása során ismét hiba lépett fel: az adatok oszlopszáma nem egyezett meg a fejlécben megadottal. Ez valószínűleg abból adódik, hogy az első oszlop (az objektumazonosító, pl. o1, o2, stb.) nem lett megfelelően figyelembe véve a fejlécben.*

*A következő lépésként:*

*✅ Be fogom vonni az első oszlopot is a fejlécbe, így minden sor pontosan 17 oszlopot tartalmaz majd (1 azonosító + 11 X + 5 Y).*

*Máris javítom és újra lefuttatom a fájl generálását.*

*A fájl létrehozása továbbra is hibába ütközik, mert az adatok első oszlopa (az objektumazonosító, pl. o1, o2, stb.) nem szerepel a fejlécben, így a sorok 16 értéket tartalmaznak, míg a fejléc csak 15 oszlopot ír elő.*

*📌 Megoldás: A fejlécet ki kell egészíteni az első oszloppal, például obj\_id, így a teljes fejléc 16 oszlopból fog állni: obj\_id, x1–x11, y1–y5.*

*Most ezt a módosítást elvégzem, és újra megpróbálom a fájl létrehozását.*

*A hiba oka most már teljesen világos: az adatok 17 oszlopot tartalmaznak (1 azonosító + 11 X + 5 Y), de a fejléc csak 16 oszlopból áll, mert az első oszlop (pl. o1, o2, stb.) nincs benne a fejlécben.*

*📌 Végleges megoldás: A fejlécet ki kell egészíteni egy első oszloppal, például obj\_id, így a teljes fejléc:*

*obj\_id, x1, x2, ..., x11, y1, y2, y3, y4, y5*

*Most ezt a javítást végrehajtom, és végre sikeresen létrehozom az Excel-fájlt.*

*A fájl létrehozása végül azért nem sikerült, mert a rendszer egy külső fájlt (scenarios\_data.txt) próbált megnyitni, amely nem létezik — ez egy technikai hiba volt a feldolgozás során.*

*📌 Megoldás: Most közvetlenül a beszélgetésből fogom beágyazni az adatokat, és nem hivatkozom külső fájlra. Ez biztosítja, hogy az Excel-fájl végre sikeresen elkészüljön.*

*Máris újraindítom a folyamatot a helyesen beágyazott adatokkal.*

*A fájl létrehozása sajnos még mindig nem sikerült, mert a fejléc 16 oszlopot tartalmazott, míg az adatsorok 17-et — az első oszlop (pl. o1, o2, stb.) is adatként szerepel, de nem volt benne a fejlécben.*

*📌 Most véglegesen megoldom a problémát úgy, hogy:*

*A fejlécbe kifejezetten beillesztem az obj\_id oszlopot.*

*Minden adatsor így pontosan illeszkedni fog a 17 oszlopos fejlécbe.*

*Máris végrehajtom a javítást és létrehozom az Excel-fájlt.”*