# Táplálkozási tanácsadó szimulátor fejlesztése, avagy modellezési stratégiák összehasonlító elemzése

Pitlik László, Ruff Ferenc

[ÖSSZEGZÉS 2](#_Toc294969401)

[Bevezetés 3](#_Toc294969402)

[Előzmények és célok 3](#_Toc294969403)

[A probléma 3](#_Toc294969404)

[Operatívan megválaszolandó kérdések katalógusa 4](#_Toc294969405)

[Szakirodalmi háttér 4](#_Toc294969406)

[Társadalmi vonatkozások 4](#_Toc294969407)

[Módszertani előzmények 5](#_Toc294969408)

[Élettani kutatások részeredményei 5](#_Toc294969409)

[Elvárás-katalógus 6](#_Toc294969410)

[Modellezési tapasztalatok 16](#_Toc294969411)

[Magyarázó modell többszörös lineáris regresszió ill. másodfokú polinomiális regresszió felhasználásával 16](#_Toc294969412)

[Adatbázis 16](#_Toc294969413)

[A regressziós vizsgálatok célja 16](#_Toc294969414)

[Alapmodell (M1) - többszörös lineáris regressziós modell 16](#_Toc294969415)

[Második modell (M2) - többszörös lineáris regressziós modell 18](#_Toc294969416)

[További háttérvizsgálatok 20](#_Toc294969417)

[A modell kiválasztása 20](#_Toc294969418)

[A változók rangsorolása 20](#_Toc294969419)

[A kapott modell értelmezése 21](#_Toc294969420)

[Lineáris részeredmények összegzése 22](#_Toc294969421)

[Harmadik modell (M3) - polinomiális regressziós (másodfokú) modell 23](#_Toc294969422)

[További vizsgálatok 24](#_Toc294969423)

[A modell értelmezése 24](#_Toc294969424)

[A lokális maximumok értelmezése 25](#_Toc294969425)

[A lokális minimumok értelmezése 25](#_Toc294969426)

[Negyedik modell (M4) - polinomiális regressziós (másodfokú) modell 25](#_Toc294969427)

[További vizsgálatok 26](#_Toc294969428)

[A lokális maximumok értelmezése 26](#_Toc294969429)

[A lokális minimumok értelmezése 26](#_Toc294969430)

[A változók sorrendje 26](#_Toc294969431)

[Genetikai potenciál számítása 26](#_Toc294969432)

[A modellek számszerű összehasonlítása 28](#_Toc294969433)

[Következtetések 28](#_Toc294969434)

[Hasonlóságelemzés 29](#_Toc294969435)

[Monoton additív modellek 30](#_Toc294969436)

[Inverz modellek üzenete 30](#_Toc294969437)

[Másodlagos monoton irányok 31](#_Toc294969438)

[Nem monoton additív modellek 32](#_Toc294969439)

[Szekunder adatos additív modellek 34](#_Toc294969440)

[Multiplikatív modellek 34](#_Toc294969441)

[A hasonlóságelemzési modellek együttes értékelése 36](#_Toc294969442)

[WizWhy 36](#_Toc294969443)

[Vita, következtetések 36](#_Toc294969444)

[Statisztikai függvényillesztés 36](#_Toc294969445)

[Hasonlóságelemzés 36](#_Toc294969446)

[Eltérő modellezési filozófiák kölcsönhatásai 37](#_Toc294969447)

[Összefoglalás, ajánlások 37](#_Toc294969448)

[Irodalomjegyzék 38](#_Toc294969449)

## ÖSSZEGZÉS

A humán só-fogyasztás (NaCl) mértékének a várható élettartamra gyakorolt hatását vizsgáltuk többféle matematikai modell alapján. Megállapítottuk, hogy a két jelenség között egy instabil optimum-hatás tételezhető fel, s ez az állami prevenciós stratégia helyességét támasztja alá - további vizsgálatokat elvárva. A legjobb modell keresése során a közismert matematikai-statisztikai mutatószámrendszeren túlmenően kiegészítő jelleggel logikai szempontrendszer is kialakításra került. Egyes logikai elvárások csak hasonlóságelemzés keretében vizsgálható. A legjobb modell meghatározása ’n’ dimenziós értékelési rendszerben önmagában is egy hasonlóságelemzést igényel, mely értékelési rendszer nem várja el a szignifikáns jelző bevezetését. A KO-feltételek alapján legjobb modellről, azaz szimulátor-építésre alkalmas, egységes szerkezetbe foglalható, rendszerszintű tudásról a vizsgált adatok alapján még nem beszélhetünk.

## Bevezetés

A következő tanulmány egy konkrét, társadalmilag releváns döntési helyzet alapján egy általános jellegű biomatematikai kérdést tárgyal új (minőségbiztosítást jelentő, tudástranszfert támogató) módszertani elvek alapján.

### Előzmények és célok

A témaválasztást egy ÁNTSZ/OÉTI & SZIE kooperációs javaslat alapozta meg 2010 elején, mely a kaposvári Alkalmazott Informatika konferencia után fogalmazódott meg [B1]. Jelen tanulmány a következőket tárgyalja: Az 1963 és 2003 közötti, 10 éves ritmusú, közhasznúan rendelkezésre álló adat-felvételezés [vö. B2] 17 EU-tagország tekintetében megadta az átlagos tápanyag-beviteli adatokat (Xi), ill. az általános demográfiai adatokkal kapcsolatos várható élettartam (Y) megoszlását országonként és évente. Az összefüggések alapján az vizsgálatok alapkutatási rétegében modellezési módszerek összehasonlítására kerül sor. A modell-versenyben a regressziós függvényillesztés, a hasonlóságelemzés [B3] és a döntési erdőként leírható WizWhy keretrendszer [B4] kerül összevetésre. Az alkalmazott kutatási rétegben az életforma-tanácsadás érdekében létrehozott (vö. dietetikai) szimulátor tanulságai kerülnek értelmezésre (pl. *Milyen tápanyag hatása milyen hatásmechanizmus során, milyen mértékű*?). A tudományos ismeretterjesztő rétegben a genetikai potenciál (a várható maximális emberi élettartam) mértéke kerül levezetésre.

A kutatás céljai: az eltérő modellezési keretrendszerek által szolgáltatott részeredmények alapján a legvalószínűbb (leginkább konzisztens) modell-logika levezetése egy robosztus szimulátor kinyerése érdekében, ill. ennek alapján az egyes hatásmechanizmusok mennyiségi jellemzőinek értelmezése azért, hogy állami megelőzési (prevenciós) programok hatástanulmányaihoz nyújtson hasznos segédletet.

### A probléma

Az alábbi tanulmány egy univerzális matematikai problémát mutat be, részben a könnyebb érthetőség, részben a szakmapolitikai aktualitás kedvéért egy konkrét (biológiai/élettani) döntési helyzetben. Ez a döntési helyzet nem más, mint elrendelhető-e tudományos hitelességgel egy állami só-prevenciós stratégiai, azaz: *Mennyi só (NaCl) bevitele (élelmiszer formájában), miként hat a várható élettartamra (az Olvasóéra és a szerzőkére egyaránt)?* Más megfogalmazásban: *A só-bevitel növelése egyenes vagy fordított arányban, vagy optimum-jelleg szerint hat-e az élettartamra a társadalom szintjén?* (A só hiánynak/többletnek tulajdonított betegségek az élettartam csökkenését jelentik formális logikai alapon. Így minden olyan szakirodalmi utalás, mely a só-bevitel szintjét valamilyen betegséggel hozza összefüggésbe, egyben a várható élettartamot is közvetlenül érinti.)

A ceteris paribus (egyetlen tényező kivételével a többi változatlansága melletti vizsgált) hatásmechanizmusok [B5] felismerésének kérdése egy általános érvényű biomatematikai/biometriai kérdés. S egyben egy univerzális matematikai/filozófiai probléma is. Az univerzális probléma nem más, mint hogy: *Két modell közül melyiket illik jobbnak tartani a másiknál, a győztesnek ítélt modell minden résztulajdonságát döntések alapjaként tekintve?* (vö. céltalanság tétele [B6])

### Operatívan megválaszolandó kérdések katalógusa

* *Létrehozható-e olyan modell-minősítő rendszer, mely tetszőlegesen sok szempont egységes keretbe foglalása alapján képes valamely jelenség-pár közötti összefüggés formájára vonatkozóan hiteles (döntés megalapozására alkalmas) megállapításra jutni, vagy a döntésképtelenséget kijelenteni?* (Feltételezetten helyes válasz: **Igen**. Elméletileg alternatív válasz: **nem**, noha nemleges válasz esetén felmerül az eddigi összes nemzetközi és hazai egészségpolitikai döntés szakmai megkérdőjelezhetősége.)
* *Elegendő-e ezen sok szempontos értékelési rendszer kialakításához csak az eddig közismert matematikai-statisztikai mutatószámrendszer?* (Feltételezetten helyes válasz: **Nem**. Elméletileg alternatív válasz: **igen**, de vajon ezen esetben: *Miért nem jött létre évszázadok óta olyan szakértői rendszer, mely a vizsgálandó jelenségre vonatkozó összes szabványosítható körülmény ismeretében világos értékelési eljárást jelölt volna ki az alkalmazandó eljárások, ellenőrzési aspektusok tekintetében a teljes szakmai közösség többségi egyetértése mellett evolúciós jellegű önfejlődést mutatva az idő múlásával?*
* *Legitimnek tűnik-e a nemzetközi tendenciákat követő, érvényben lévő hazai életvezetési stratégia a só-bevitel csökkentésének szorgalmazására?* (Feltételezetten helyes válasz: **Nem állapítható meg a vizsgált adatok alapján**. Elméletileg alternatív válasz: **igen**, ill. **nem**, vagyis egyértelmű bizonyítékláncolat tárható fel pro vagy kontra.)

## Szakirodalmi háttér

Mint az már a bevezetésben, ill. a szakirodalom szerkezetén keresztül is érzékelhető, a szakirodalmi háttérként sok esetben internetes hivatkozások kerültek feltárásra, melyek értékéért a tanulmány szerzői vállalják a felelősséget. A XXI. században a klasszikus szerkesztőségek minőségteremtő munkája és az ehhez szükséges (késleltetési) idő ugyanis nem feltétlenül állnak arányban egymással. Tudatos értékvállalásként kijelenthető: Nem a megjelenés helye, formája, módja, hanem egyedül és kizárólag a megjelent tartalom teremt elsődlegesen értéket.

### Társadalmi vonatkozások

A szakirodalmi háttér több rétegű: egyrészt létezik magának az alkalmazási élethelyzetnek (vö. táplálkozási tanácsadó szimulátor) szakirodalmi előzménye. Ha a teljesség igénye nélkül elolvassuk csak az alábbi három dokumentumot [B7, B8, B9], akkor (úgy átlagos állampolgárként, mint kutatóként) joggal merül fel a kérdés: *Szabad-e ilyen mértékben egymásnak ellentmondó nézeteknek egyáltalán léteznie egy futó prevenciós stratégia esetében?* (a kibernetikus jogállamisággal kacérkodó XXI. században. Az ugyanis nem a sajtó- és gondolatszabadság kérdése, hogy jogában álljon bárkinek is adott pillanatban bizonyíthatóan tévesnek nevezhető közléseket tenni.) Tudományos értelemben a feladat tehát egyértelmű: *Milyen bizonyítási eljárás keretében lehet a két vitázó álláspont közül bármelyik igazságát kijelenteni, ill. ki lehet-e jelenteni bármit is „megfelelő” bizonyossággal? Mi ennek a bizonyosságnak a mértékegysége? Milyen bizonyítási eljárás keretében állítható elő bármilyen tartalmú, hiteles szakvélemény?*

Ha ugyanis egy ceteris paribus alakzat mellett a tudományos igazolást le lehet tenni, akkor ez egy kibernetikus jogállam esetében valóban prevenciós stratégiák alapja lehet a társadalmi sokaság szintjén (de az egyéni sokszínűség még ekkor is „nyitott” kérdés: vö. Liebig-féle minimum elv egyed és populáció szintű „ellentmondásosságának” feloldása: pl. [B10].

Ha azonban kellő bizonyossággal nem vélelmezhető egyik ceteris paribus alakzat sem, akkor bármelyiket is állítjuk egy prevenciós stratégia középpontjába, jogi értelemben pl. a gondatlan veszélyeztetés gyanúja vetődhet fel, közpénzen finanszírozva magát a károkozást.

Bármennyire patetikusan is hangzik: Így válik a biomatematika/biometria a kibernetikus jogállam még meg nem írt alkotmányának központi kérdésévé.

Biomatematika [B11] és biometria alatt elsődlegesen a közismert matematikai-statisztikai ismeretanyagot értik az Olvasók [vö. B21]. A tanulmány egyik újszerűsége abban áll:

*Hogyan lehet egy párhuzamos módszertani megközelítést jelentő eljárást, a konzisztencia-orientált hasonlóságelemzés logikai rétegeit ötvözni az ismert függvényillesztési eljárásokkal annak érdekében, hogy a leghitelesebb ceteris paribus alakzatok felderítése minél inkább automatizálható legyen?*

### Módszertani előzmények

A matematikai-statisztika hagyományos ismeretanyagának rövid felvázolását a szerzők itt és most nem tekintik feladatuknak. Az elvárás-katalógus című fejezetben azonban részletes újraértelmezése következik a modellek jóságát megítélni engedő egyes mutatószámoknak, annak érdekében, hogy a keresési stratégiákra alapozó modellezések szempontjából is nagyító alá kerüljenek triviálisnak tűnő, berögzült gondolati panelek valódi hatásai.

A módszertani háttér másik rétege a biológiai kérdések matematikai vizsgálatának alternatív megközelítését érinti: Jelen tanulmány módszertani előzményei visszanyúlnak az orvosi szakértői rendszerek holisztikus stabilitásának hasonlóságelemzés keretében történő értelmezésére (vö. [B12]), éppen úgy, mint a tér/idő modellezésének univerzális megközelítésére (vö. [B13]).

A hasonlóságelemzés átfogó értelmezését a <http://miau.gau.hu/myx-free> szolgáltatás keretében lehet több nyelven is megtalálni, ill. a Magyar Internetes Alkalmazott/Agrár Informatikai Újság vezércikk sorozata ad sokszínű áttekintést a hasonlóságelemzésről (vö. [B14]).

A hasonlóságelemzés lényege, hogy az eddig kevéssé vizsgált lépcsős függvények kombinációira támaszkodva eddig meg nem válaszolt kérdéseket (pl. *Klasszifikációs, csoportosítás, klaszterezési problémák esetén létezhet-e az objektumok azonossága? Miként fejezhető ki egy objektív ár/teljesítmény, tény-norma viszony n-dimenziós benchmarking/SWOT erőtérben? Miként hasznosítható az a tanulási input-mintázatok inverz alakja a módszertani kockázat, ill. érzékenység fogalmának újraértelmezésére? Milyen ceteris paribus alakzatok tárhatók fel automatikusan egy adott tanulási minta változói között eltérő keretfeltételek esetén?*) tegyen értelmezhetővé.

### Élettani kutatások részeredményei

Ezen szakirodalmi réteg a számára külön nem kerül kifejtésre, mert ez a hagyományos biometriai megközelítés szerves részeként kerül majd bemutatása a modellezési tapasztalatok fejezet statisztikai függvényillesztés című alfejezetében.

## Elvárás-katalógus

A matematikai-statisztika eszköztárának ismerői, oktatói, alkalmazói sem szabad, hogy meghökkenjenek azon tapasztalati tényen, miszerint egy adott jelenség másik jelenségre gyakorolt hatásának feltárására - a rendelkezésre álló módszertani paletta tetszőlegesen kiragadott eszközkészlete alapján - nagy valószínűséggel bármely erre kiképzett Hallgató minden különösebb vonakodás nélkül formál a mindenkori szakmai szempontok alapján jelesre értékelhető „megoldást” rövid idő alatt könnyen elérhető adatvagyon és kiforrott szoftvertámogatás mellett.

Ez azt jelenti, hogy a jogi, politikai, társadalmi felelősségvállalástól leválasztva, a módszertani tudás-fragmentumok ismeretében látszólag problémamentesen lehetne nyilatkoztatni erre a célra kiképzett szakembereket arról: *Vajon a só-bevitel csökkentésén vagy növelésén keresztül lehetne magasabb átlagéletkort elérni?*

A tanulmány írása kapcsán ad hoc jelleggel meginterjúvolt, elismert és nagy tapasztalattal rendelkező élettani szakértők egy része azonban azonnal és mereven elzárkózott az itt feldolgozott adatvagyon alapján a kérdés megválaszolásától.

A szakértői attitűdöket érintő két tapasztalati tény egymás formális ellentettje (rutin feladat vs. lehetetlenség). A mindenkori egészségpolitika, ill. a tudástranszferben résztvevő egyéb szereplők azonban mégis csak ilyen jellegű, minőségű és mennyiségű adatvagyonok alapján kell, hogy döntésekre jussanak. Tehát módszertani szempontból szükséges annak áttekintése: *Milyen minőségbiztosítási értékelési rendszer keretében képzelhető el az eddig felhalmozott módszertani tudásrétegek egymásra épülő hasznosítása egy maximálisan hiteles döntés megalapozása érdekében?*

Az alábbi nem teljes körű áttekintés célja, hogy a közismert matematikai-statisztikai fogalmakat a minőségbiztosítás és a keresésvezérlés eszközeiként is megvizsgálja, ezzel alapozva meg a hasonlóságelemzés bevonásának szükségszerűségét a modellek értékelésébe, a ceteris paribus hatásmechanizmusok feltárásába:

Egy logikai egységben először a mutatószám neve, majd ennek rövid leírása szerepel. Ezt követi annak meghatározása (vö. STAT= statisztikai-matematika, HE=hasonlóságelemzés), hogy *a mutatószám számítható-e hasonlóságelemzési modellek esetén alapvetően a becsült és a tényértékek birtokában (vagyis mely mutatószámokon keresztül lehet a hagyományos modellezési logika és a hasonlóságelemzés között hidat verni)?* Majd egy fajta státuszváltozóként deklarálásra kerül, *vajon az adott mutatószám alkalmas-e (s ha igen milyen formában) KO-feltételként való hasznosításra a modellek értékelésének folyamatában?* A legutolsó pontban kifejtésre kerül, vajon milyen addicionális, alternatív értelmet nyerhet az adott mutatószám, ha egy modell-kereső/értékelő rendszer részeként akarjuk értelmezni.

A mutatószámok/elvárások két csoportban kerülnek bemutatásra: az első blokkban a közismert statisztikai jellegűek, míg a második blokkban azon, logikai jellegű mutatószámok, melyek alapvetően a hasonlóságelemzési modellezés konzisztencia-vizsgálatait során kerülnek alkalmazásra, de bármely más modell-logika esetén is igaznak kell lenniük.

A katalógus elemeinek bemutatása előtt szükséges utalni arra, hogy a céltalanság tétele értelmében a modellek közötti választás matematikai alapjai kiforratlanok. Ennek első vetülete, hogy az egyes korrelációs koefficiensek (vö. kontingencia, pearson, numerikus korreláció – vö. [B15]) egymástól bizonyos „térben” teljesen függetlenek. Az n-rétegű konzisztencia fogalma jelenti a kitörést a modell-értékelés jelenlegi zsákutcájából: (vö. [B16]).

**A modellek értékelése tetszőlegesen sok mutatószám alapján nem más, mint egy bírósági tárgyalás a modellező ügyész és a modellező védő által feltárt bizonyítékok alapján arról, vajon egy-egy modell alkalmas-e arra, hogy alkalmazásra kerüljön adott kérdés megválaszolására…**

Az elvárás-katalógus elemei – statisztikai mutatószámok:

|  |  |
| --- | --- |
| Elnevezés: | AIC (Akaike information criterion) |
| Leírás: | <http://en.wikipedia.org/wiki/Akaike_information_criterion> |
| Jelleg: | STAT & HE |
| Típus: | sorba rendezésre alkalmas, KO-feltételként nem alkalmazható |
| Értelmezés: | Az R2 alternatívája, más módon méri a becslési pontatlanság modell szintű összességét. Ennek alapján előfordulhat két modell összehasonlítása során, hogy az R2 növekedése az egyik modell esetén AIC értéknövekedéssel = modellérték-csökkenéssel jár együtt. Alkalmas a túltanulás kezelésére. |

|  |  |
| --- | --- |
| Elnevezés: | Adjusted R-squared |
| Leírás: | <http://en.wikipedia.org/wiki/Coefficient_of_determination> |
| Jelleg: | STAT & HE |
| Típus: | sorba rendezésre alkalmas, KO-feltételként nem alkalmazható |
| Értelmezés: | A változók számára jutó hiba fogalma lehetővé teszi a modellezés értékelésében a túltanulás elleni küzdelmet: a változók számának növelésével a modellek egyre könnyebben képesek magasabb becslési pontosság elérésére, azonban a mutatószám által sugalmazásra kerül egy ekvivalencia: mennyi változószám-növekedés mennyi pontosságnövekedést illene, hogy magával hozzon: 1-((1-R^2)\*(N-1)/(N-k-1)), ahol ’N’ az objektumok száma, ’k’ a bevont változók száma. Amennyiben kísérleti jelleggel manuálisan konstansként határozzuk meg az egy változószám-növekedéstől elvárt R^2 változást, akkor ennek hatása a fenti képlet kimenetére nem konstans: 0.00126-ig növekvő mértékben degresszív, míg 0.00127 esetén már növekvő mértékben progresszív: vagyis a változószám/szabadságfokok változásakor nem mindegy, hogy milyen szinten nőnek egy egységgel. Ez az átcsapás a mutatószám értékelő erejét veszélyezteti…      K1. ábra: A változószám-változás szimulált hatásai (Forrás: saját számítások)  A K1. ábra a jelenlegi tanulási minta kondícióira érvényesítve mutatja be a számítást.  A változók számának növelését büntető elv emellett ütközik azon filozófiai elvárással: miszerint minden mindennel összefügg, ellenben erősíti az Einstein nevéhez kötődő elvet: vö. az egyszerűbb a jobb. A modellek komplexitásának büntetése tehát önmagában nem alkalmas a valódi túltanulási effektusok (látszat-korrelációk) kiszűrésére. A túl sok magyarázó változó csökkenti a szabadságfokot (a megfigyelések és a becsülni kívánt paraméterek számának különbségét), s ezáltal nem engedi meg a becslés statisztikai tulajdonságainak érvényesülését. Ezt képes pótolni, kiegészíteni a többrétegű konzisztencia vizsgálat. |

|  |  |
| --- | --- |
| Elnevezés: | Multiple R-squared |
| Leírás: | <http://en.wikipedia.org/wiki/Coefficient_of_determination> |
| Jelleg: | STAT & HE |
| Típus: | KO-feltételnek alkalmas: pl. a kibernetikus jogállam szempontjából az életformát befolyásoló prevenciós intézkedések esetén elvárható egy önkényesen deklarált érték feletti modelljóság: (pl. 0.8) |
| Értelmezés: | Azt mutatja, hogy mekkora hányadban magyarázzák a független változók a függő változó teljes eltérés négyzetösszegét. A mutató nem alkalmas a túltanulás feltárására, leleplezésére. Célfüggvény-értékként kifejezetten túltanulásra csábít. |

|  |  |
| --- | --- |
| Elnevezés: | Residual standard error ( or standard error of estimate) |
| Leírás: | <http://en.wikipedia.org/wiki/Mean_squared_error> |
| Jelleg: | STAT & HE |
| Típus: | sorba rendezésre alkalmas, KO-feltételként nem alkalmazható |
| Értelmezés: | A tény és a becslés eltérésének értékelésekor semmilyen mutató nem segíti feloldani azt az alapproblémát, vajon egy nagy hiba és sok apró, vagy sok átlagos hiba esetén magyaráztuk a modellel a valóságot helyesebben? Vagyis túltanultnak számít-e az a modell, mely olyan nagy eltéréseket próbál megmagyarázni rendelkezésre álló adatokból, melyeknek nem is ezen függetlenváltozók az okai? |

|  |  |
| --- | --- |
| Elnevezés: | F-statistic: |
| Leírás: | <http://en.wikipedia.org/wiki/F-test> |
| Jelleg: | STAT |
| Típus: | KO-feltételnek alkalmas: pl. a kibernetikus jogállam szempontjából az életformát befolyásoló prevenciós intézkedések esetén elvárható egy önkényesen deklarált érték feletti modell-helyesség… |
| Értelmezés: | A mutatószám megadja, hogy a vizsgált változók feltárt együtthatói közül legalább egy szignifikánsan különbözik-e a nullától. |

|  |  |
| --- | --- |
| Elnevezés: | variance analysis: residuals to fitted values  standardized residuals to fitted values (scale location)  standardized residuals to theoretical quantiles(normal Q\_Q)  standardized residuals to leverage |
| Leírás: | <http://en.wikipedia.org/wiki/Q-Q_plot>  <http://www.sciencedirect.com/science/article/pii/016794739190059B>  <http://stat.ethz.ch/R-manual/R-patched/library/stats/html/plot.lm.html> |
| Jelleg: | STAT & HE |
| Típus: | KO-feltételnek alkalmas, amennyiben a lineáris modell felállíthatóságát akarja valaki mérni ezen keresztül |
| Értelmezés: | Ebben az esetben a lineáris modell felállíthatóságának feltételei kerülnek vizsgálatra. Ez tehát nem modell-jóságot mérő mutatószám. A modellek végső összevetésében nem lehet már szerepük, mert: ha nem teljesíti egy modell ezen elvárásokat, szóba sem jöhet, mint helyesen illesztett lineáris modell. Ez szakirodalmi megközelítés szerint csak a regresszióra vonatkozik. De vajon miért nem érvényes ez a mutató egy hasonlóságelemzési modell esetén? Hiszen a linearizálhatóság, ill. monotonizálhatóság nem a modell, hanem a kiindulási tanulási minta tulajdonsága.  Az ábrák (vö. 1. ábra) kiértékelésének értelmező (automatizálható) szabályrendszere nem ismert (vö. képfelismerés)! |

|  |  |
| --- | --- |
| Elnevezés: | VIF-analysis |
| Leírás: | <http://en.wikipedia.org/wiki/Multicollinearity> |
| Jelleg: | STAT & HE |
| Típus: | KO-feltételnek nem alkalmas |
| Értelmezés: | A modellek módosítása során (pl. stepwise eljárás) az egymással összefüggő változók kiesnek. A VIF-mutató a modellbe bevont változók közötti összefüggést méri, vagyis az számít jobb modellnek, mely esetében a bevont változók között minimális az összefüggés. Ha azonban egy sorozatnyi, logikai szinten zajnak vélhető, de egymással kapcsolatban nem álló változót is belekeverünk az X-tömbe és emellett egymással szorosabban is összefüggő változók is vannak a bevonhatók között, akkor mindezek értelmében az lesz a helyesebb modell, ami logikai szinten lehetetlen, azaz logikai zajokat/zavaró tényezőket vonna be az eljárás. Látszólag igaz, hogy: ha az ún. zajoknak nincs kapcsolata az Y-nal, akkor nem vonja be őket, ha van, akkor pedig már nem zajok. De sajnos ismert a látszatkorreláció jelensége is (pl. az ország felett elrepülő verebek éves száma és pl. a munkanélküliség is alakulhat hasonló lefutás mellett, az egyikkel még sem illik a GDP-becslés során dolgozni). Ezt a modellezés önmagában nem tudja kivédeni. Itt már be kell avatkoznia a modellezőnek magának: pl. már az adatok összegyűjtésekor.  Ez a fentiek alapján azonban tudatos és/vagy gondatlan manipulációk alapja lehet, hiszen a változók egymás függvényében hatnak, vagy nem hatnak szignifikánsan és adott előjellel/módon.  A hasonlóságelemzés a rangsorszámokra konvertálás kapcsán ismeri és elismeri a változócsoportok fogalmát, melyek azonos/hasonló hatásmechanizmust írnak le. Ebben az esetben már nem multikollinearitásról beszélünk, hanem arról, hogy egymástól elnevezésükben független, de a vizsgált mintában hasonlóan megjelenő mechanizmusok létezhetnek, melyek hatása a becslésekre azonos logikájú, egymástól szét nem választható.  Itt említendő hasonlóságelemzési specialitás az is, hogy egy diszkrét, kombinatorikai modellben a becslés fogalma is kombinatorikai jellegű: egy adott input-adatsor minden egyes változója tekintetében 2-2 szomszédos, leghasonlóbb megfigyelési egységgel (lépcsőfokkal) írható le általában. Ebben a pillanatban a becslés 2^k darab egyedi becslés eloszlása (Y-tengelyre vetített előfordulási gyakoriságai) alapján értelmezhető csak helyesen, ahol ’k’ a változók száma. A hasonlóságelemzés tehát már önmagában a becslés fogalmában is benne hordozza annak lehetőségét, hogy a becslést alkotó hasonlóságok nem normál, vagy egyenletes megoszlásúak, hanem pl. két szélsőértéket mutatnak fel, ami adott élethelyzetben a becslés alapját adó modell helyességétől függetlenül is képes a konkrét egyedi becslés legitimitását korlátozni. |

|  |  |
| --- | --- |
| Elnevezés: | t-test |
| Leírás: | <http://www.weibull.com/DOEWeb/hypothesis_tests_in_multiple_linear_regression.htm> |
| Jelleg: | STAT |
| Típus: | KO-feltételnek nem alkalmas mutatószám |
| Értelmezés: | A kockázatot mérő mutatószám azt adja meg, hogy az egyes változók koefficiense 0-nak tekinthető-e (azaz nincs hatása), vagy pedig nem tekinthető 0-nak. Itt is felmerül annak problémája, hogy adott pillanatban rendelkezésre álló változócsoportból (jelen esetben 15 darab) egy automatikus modellépítés esetén automatikusan beválasztott és szignifikánsnak tűnő változók kihagyása után az új modellben, a szűkített változócsoport alapján ismét lehetnek szignifikáns változók egy „gyengébb” modellben.  Felmerül a kombinatorikai kérdés: ha ’k’ változóból minden létező variánst (’k’ alatt az ’i’, ahol ’i’ = 1, …, ’k’) kiválasztunk, vajon ezen modell-sokaság alapján (vö. modellezési érzékenységvizsgálat) nyerhetők-e ki olyan KO-feltételek, rangsorolási, modell-kockázati szempontok, melyek a legjobb modell kiválasztásához hozzájárulhatnak? (Ezen kérdés vizsgálata meghaladja a tanulmány tervezett kereteit). Ez a logikai elvárások között önálló elvárásként jelenik meg. |

|  |  |
| --- | --- |
| Elnevezés: | component + residual plots (for each attribute) |
| Leírás: | <http://en.wikipedia.org/wiki/Partial_residual_plot> |
| Jelleg: | STAT (& HE, amennyiben együttható \* bemenő adat = lépcsőfok) |
| Típus: | KO-feltételnek nem alkalmas |
| Értelmezés: | Az ábrák (vö. 3. ábra) kiértékelésének értelmező (automatizálható) szabályrendszere nem ismert (vö. képfelismerés)! |

|  |  |
| --- | --- |
| Elnevezés: | Variable relative importancs: img |
| Leírás: | <http://www.jstatsoft.org/v17/i01> |
| Jelleg: | STAT |
| Típus: | KO-feltételnek nem alkalmas |
| Értelmezés: | A modellbe bevont változók fontossági sorrendjét adja meg a mutatószám. |

|  |  |
| --- | --- |
| Elnevezés: | Variable relative importancs: pmvd |
| Leírás: | <http://www.jstatsoft.org/v17/i01> |
| Jelleg: | STAT |
| Típus: | KO-feltételnek nem alkalmas |
| Értelmezés: | A modellbe bevont változók fontossági sorrendjét adja meg a mutatószám. |

Logikai elvárások:

|  |  |
| --- | --- |
| Elnevezés: | A szisztematikus alul/fölül-becslés kizárása |
| Leírás, indoklás: | A becslések összege legyen egyenlő a tények összegével, hiszen ennek hiányában szisztematikus alul/fölül-becslésről kellene beszélni. |
| Jelleg: | STAT & HE |
| Típus: | KO-feltétel egyedi modellekre |
| Értelmezés: | Az LP-alapú közelítő megoldásokban ezen elvárás teljesülését külön ellenőrizni kell. A túltanulás detektálása kapcsán a KO-feltétel feloldása érdekes következményekre vezethet: nem logikátlan ugyanis az sem, ha a KO-feltételt úgy fogalmazzuk át, hogy minél több objektum esetén legyen igaz, hogy a becslések és a tények összege azonos. Ebben az esetben tetten érhető, ha egy magyarázó változó körtől független jelenség felelt egy adott output-értékért: pl. kukorica terméseredményét leíró modellbe egy olyan tábla adatai is belekeverednek, mely esetén a termés jelentős részét ellopták, ill. pl. meteor-becsapódás miatt megsemmisült. |

|  |  |
| --- | --- |
| Elnevezés: | A változók fontosságának stabilitása |
| Leírás, indoklás: | A változók logikai fontossági sorrendje eltérő modellek esetén legyen minél inkább hasonló, hiszen a modellek egymással szembeállítása érzékenységi vizsgálatra alkalmas elvárás. Minél kevésbé érzékeny egy modell-tulajdonság az egyes modellezési elvárásokra, annál stabilabban lehet kijelenteni a legjobb modell adott tulajdonságáról, hogy az úgy a leghitelesebb, ahogy az a legjobbnak tűnő modellben fellelhető. |
| Jelleg: | STAT & HE |
| Típus: | KO-feltétel modellek összehasonlításakor (azaz magas instabilitás esetén semmilyen modell nem fogadható el megalapozottan) |
| Értelmezés: | Kockázatot mérő mutatószám: A hasonlóság esetében a fontosság nem más, mint a becslések attribútumonkénti összegeinek %-os megoszlása, ahol az attribútumonkénti minimumokkal korrigálni kell a számítás során az LP direkt megoldását annak érdekében, hogy a közelítő megoldási technika számára kialakított mozgástér eltolás-hatását kizárjuk. |

|  |  |
| --- | --- |
| Elnevezés: | A változók ceteris paribus határmechanizmusának stabilitása |
| Leírás: | A változók ceteris paribus alakzatai eltérő modellek esetén legyenek minél inkább hasonlóak, hiszen a modellek egymással szembeállítása érzékenységi vizsgálatra alkalmas elvárás. Minél kevésbé érzékeny egy modell-tulajdonság az egyes modellezési elvárásokra, annál stabilabban lehet kijelenteni a legjobb modell adott tulajdonságáról, hogy az úgy a leghitelesebb, ahogy az a legjobbnak tűnő modellben fellelhető. |
| Jelleg: | STAT & HE |
| Típus: | KO-feltétel modellek összehasonlításakor (azaz magas instabilitás esetén semmilyen modell nem fogadható el megalapozottan) |
| Értelmezés: | Kockázatot mérő mutatószám: Amennyiben feltárhatók olyan modellek, melyek a hipotetikus, ill. más modellezési módszertanok által hitelesnek tűnő irányokkal szemben alternatív ceteris paribus értelmezéseknek adnak teret kellő hitelesség mellett, ez egyben felfogható ellenbizonyításnak is.  A hasonlóságelemzés klaszterezési célra is hasznosítható eljárás (vö. Y0, azaz [B20]), amennyiben kereshető olyan lépcsős függvény, mely lépcsőfokai alapján bármilyen csoportosítás kizárható (vö. matematikai pozitív diszkrimináció).  Amennyiben egy adott modellben a változók ceteris paribus alakzata nem lineáris, s a szélsőérték kívül esik a vizsgálandó (értelmezési) intervallumon, úgy a ceteris paribus alakzat monoton összefüggésként is felfogható a regressziós függvényillesztés során.  A hasonlóságelemzésben ezen jelenség párja a lépcsők regressziós összefüggésként való leírása már az irányított keresés során. Ebben az esetben az egyébként monoton lépcsők nem monoton egyenletekkel is közelíthetők, különösen akkor, ha ez rendszerelméletileg védhető feltételezés. |

|  |  |
| --- | --- |
| Elnevezés: | A változók másodlagos ceteris paribus határmechanizmusának stabilitása |
| Leírás: | A változók másodlagos ceteris paribus alakzatai eltérő modellek esetén legyenek minél inkább hasonlóak, hiszen a modellek egymással szembeállítása érzékenységi vizsgálatra alkalmas elvárás. Minél kevésbé érzékeny egy modell-tulajdonság az egyes modellezési elvárásokra, annál stabilabban lehet kijelenteni a legjobb modell adott tulajdonságáról, hogy az úgy a leghitelesebb, ahogy az a legjobbnak tűnő modellben fellelhető. |
| Jelleg: | HE |
| Típus: | KO-feltétel modellek összehasonlításakor (azaz magas instabilitás esetén semmilyen modell nem fogadható el megalapozottan) |
| Értelmezés: | Iránykereső modellezéskor (COCO\_MCM) másodlagos, monoton ceteris paribus irány levezetése lehetséges minden egyes attribútumra. A másodlagos irány az MCM-lépcsők meredeksége alapján (regressziós logika szerint) határozható meg polinomizálódó lépcsők esetén is. |

|  |  |
| --- | --- |
| Elnevezés: | A rendszerelméleti keretfeltételek betartása |
| Leírás: | A nem monoton összefüggések esetén a lokális minimumok kizárhatók, hiszen az élettani mutatók esetében köztudott, hogy bármilyen a hiánya (minimális input érték), ill. bármilyen változó esetén a túlzott adagolás (vö. mérgezés) a következményváltozóra (várható élettartamra) negatívan hat. |
| Jelleg: | STAT & HE |
| Típus: | KO-feltétel egyedi modellek esetében |
| Értelmezés: | Amennyiben egy változónak ceteris paribus hatásmechanizmusa nem monoton, akkor a lokális minimum elismerése a minimális input és a maximális input abszolút minima mellett egy hullám-függvényt eredményezne, mely jelen ismereteink szerint csak akkor léphet fel, ha a vizsgált mintát (az adat-felvételezés módszertanát) nem tekinthetjük homogénnek/reprezentatívnak. Jelen esetben az EU népessége országonként és évtizedenként formálisan konszolidált/homogén az egységes adatgyűjtési módszertanok alapján. |

|  |  |
| --- | --- |
| Elnevezés: | A kísérleti részösszefüggésekkel való összhang |
| Leírás: | Egyed/sejt szintű vizsgálatok során nyert egyváltozós részösszefüggések és a többváltozós modell ceteris paribus alakzatainak összhangja legyen minél nagyobb mértékű. |
| Jelleg: | STAT & HE |
| Típus: | KO-feltétel egyedi modellek esetében |
| Értelmezés: | Elvileg fel kellene tételeznünk, hogy egy adott ceteris paribus helyzetben felállított klasszikus kísérlet (pl. só-dózis növelés hatása a sejtek életben maradásának hosszára) egyváltozós összefüggésre vonatkozó eredménye egyed/populáció szinten is igaz. Ezzel a mondattal lényegében a tanulmány alapkérdését is ignorálnánk, hiszen akkor ilyen jellegű alapkísérleteket kellene beállítani populáció szintű statisztikai adatvagyonok kiértékelése helyett. Azonban az is igaz, hogy eltérő ceteris paribus feltételek mellett (pl. homoktalaj, barna erdőtalaj, tőzegtalaj) a nitrogén terméseredményre gyakorolt hatása más-más ceteris paribus alakzat szerint detektálható: homokon a racionális értelmezési intervallumban alapvetően monoton növekvő alakzat lesz tetten érhető. Mmíg egy átlagos termőképességű talajon kialakulhat a közismert csökkenő hozadék elvéhez tartó maximum szélsőértékkel rendelkező görbe. S végül egy nitrogénben gazdag talajon bárminemű további nitrogén adat csökkenő hatással bírhat a terméseredményre. Ezen három karakterisztika egy ország tápanyag-gazdálkodási adatsorán az egyes talajok arányától függően alakul ki, s vehet fel a három közül bármelyikre hasonlító alakzatot, mely tehát az ország, mint rendszer ceteris paribus alakzata. Itt kell tehát megerősíteni, hogy jelen tanulmány az EU, mint rendszer átlagos karakterisztikáit keresi a vizsgált változók esetében, feltételezve, hogy a prevenciós intézkedések sem tudnak érdemi különbséget tenni a marketing üzenetek rész-célcsoportjai között. |

|  |  |
| --- | --- |
| Elnevezés: | Tanulási többlettudás legyen tetten érhető |
| Leírás: | A genetikai potenciál és a minimális rendszer output intervalluma legyen szélesebb, mint az ismert tanulási minta értelmezési intervalluma a következményváltozó esetén. |
| Jelleg: | STAT & HE |
| Típus: | KO-feltétel egyedi modellek esetén |
| Értelmezés: | Amennyiben egy tanulási / függvényillesztési eljárás erre nem képes, akkor az túlzottan restriktív, azaz a rendelkezésre álló adatokból nem képes a tetszőleges túltanulási szint elérésére. Ugyanis minden modellezés túltanulásra törekszik saját keretfeltételei között, lévén hibát minimalizál addig, amíg a keretfeltételek között lehetséges. Mivel a tanulási minták indirekt módon, de leírják az ideális kombinációt is, így erre vonatkozóan racionális elvárás, hogy ennek hatása a rendszer-outputra legalább akkora illik, hogy legyen, mint a már ismert rendszer-maximum. Ennek fordítottja is igaz: a leggyengébb inputkombináció, melynek következménye a legjobbhoz hasonlóan csak indirekt módon ismert, nem vezethet jobb output-értékre, mint egy kevésbé rossz, ismert input-kombináció. Jelen tanulmányban a vizsgált várható élettartamok: 80.3 és 64.8 év között ingadoztak. |

|  |  |
| --- | --- |
| Elnevezés: | Inverz modellek viselkedése legyen racionális |
| Leírás: | A tanulási minta manipulálásán keresztül kikényszerítetten, változónként ellentétes előjellel rendelkező modellek egymás inverzei, s ez az outputok mennyiségi és minőségi tükörképként való megjelenésében érhetők tetten. |
| Jelleg: | STAT & HE |
| Típus: | KO-feltétel az egyedi modellek esetén (amennyiben az inverziós hibák száma elér egy önkényesen kinevezett kritikus szintet) |
| Értelmezés: | A regressziós eljárásokban ez magától értetődő tény, amennyiben a tanulási mintát változónként saját maximumából kivonjuk. A hasonlóságelemzés esetén a lépcsők monoton, de tetszőleges lefutást képesek biztosítani, így arra is képesek, hogy a tükör-effektust feltörjék: a modellezés és az értelmezés inverz logikája szerint (vö. kettős tagadás), ha létezhet olyan paraméter-tömb (lépcsősorozat), mely esetén a hiba-minimalizálás és az inverz alakzatok szimmetriája elszakad egymástól, ott modellezési kockázatok érhetők tetten. S minél nagyobb ezen kockázatok aránya az összes objektumhoz képest, annál hiteltelenebb bármilyen modellezési eljárás eredménye. |

|  |  |
| --- | --- |
| Elnevezés: | Változók közötti kapcsolatok legyenek racionálisak |
| Leírás: | A tanulási folyamatban erre való külön kényszerítés nélkül elvárható, hogy az egymással összefüggő független változók kölcsönhatásai megfelelnek a szakirodalmi ismeretnek. |
| Jelleg: | STAT & HE |
| Típus: | KO-feltétel egyedi modellek esetében |
| Értelmezés: | A hasonlóságelemzés esetén az eredetileg független változók egymással is kapcsolatba kerülnek, amikor minden egyes független változó Y-ként kerül értelmezés a többi, fennmaradó független változó és az eredeti függőváltozó alapján. Ebben az esetben a modellező alapkérdése: Mennyinek illett volna lennie az adott független változó értékének, a függőváltozó ismert értéke alapján? Ezen n-rétegű ellenmodellek felállításakor elvárható, hogy az egymással pl. halmazelméleti szinten összefüggő jelenségek (pl. telített és telítetlen zsírsavak) ennek megfelelően reagálnak, s a legfontosabb magyarázó tényező az egyik esetében a másik tényező lesz. |

|  |  |
| --- | --- |
| Elnevezés: | A változók hatásmechanizmusainak egymástól független stabilitása |
| Leírás: | A statisztikai elvárások között is már jelzésre került, hogy tetszőleges változócsoportok alapján egy adott változó hatása, előjele, mértéke legyen minél stabilabb. |
| Jelleg: | STAT & HE |
| Típus: | KO-feltétel több modell vonatkozásában, amennyiben erre önkényesen legitim küszöbérték kerül kijelölésre jogállami keretek között |
| Értelmezés: | A többváltozós modellezési eljárások a tanulási mintát mindenkor a modell-egyenlet valamilyen definíciót jelentő hibájának minimalizálása mentén képzelik el. Ebben a logikai keretben a modell, s nem a változó a főszereplő, tehát hasonló furcsaságok alakulhatnak ki, mint a termelési függvények esetében pl. az NPK helyettesíthetőségek feltételezésekor. Élettani, sejtes kísérletekben a három anyag biokémiai logikája egészen más, lényegében nem helyettesíthető, mégis: a termelési függvények matematikai tulajdonságai azt sugallják, mintha ezek magas szinten egymást kiváltó anyagok lennének élettani szinten is, noha nincs másról szó, mint arról, hogy a marginális elemzési elvek szerint az egyedi, termésre gyakorolt hatások eredője több ponton ekvivalenciákat mutat, aminek élettanilag semmilyen értelmezhetősége nincs.  Jelen esetben: minél inkább feloldjuk a modellezés restrikcióit, pl. monotonitást, a nem nemlineáris függvények szimmetriáját, annál szabadabban lehet felismerni, hogy az egyes változók a többi jelenlététől függően jelentősen eltérő módon hat a következményváltozó magyarázata érdekében. Már a független változókra érvényesített előjel-variánsok is egészen eltérő viselkedés mintákat hívnak elő a tanulási minta alapján a változók fontosságát érintően. |

|  |  |
| --- | --- |
| Elnevezés: | Szimmetria elvárások feloldása |
| Leírás: | Az optimumtól való távolodás két lehetséges iránya jelentősen nem azonos hatásmértékkel bír a függőváltozóra. |
| Jelleg: | STAT & HE |
| Típus: | KO-feltétel több modell egymással való versenyeztetése esetén |
| Értelmezés: | A C-vitamin hiánya és túladagolás esetén a várható élettartam az optimumtól távolodva jelentősen másként viselkedik. A hiány meredek zuhanást, míg a túladagolás látszólagos határtalanságot mutat, ill. csak nagyon extrém esetekben vezet érzékelhető csökkenéshez. A közismert függvényillesztési eljárások optimum körüli szimmetria-kényszere abban az eset értelmezendő KO-feltételként, ha a regressziós eljárásokkal szemben, más, rugalmasabb keretfeltételeket mutató eljárások vélelmezni engedik a szimmetria nemlétét. |

|  |  |
| --- | --- |
| Elnevezés: | Rendszer-szélsőségek tolerálása |
| Leírás: | Amennyiben a primer megfigyeléseket kiegészítjük szekunder objektumokkal, akkor a modell ezeket legyen képes racionálisan kezelni. |
| Jelleg: | STAT & HE |
| Típus: | KO-feltétel több modell összevetése kapcsán |
| Értelmezés: | A korábbiakban is említésre került már, hogy az egyes élettani inputok teljes/relatív hiánya, ill. jelentős túladagolása egyértelműen meredek élettartam csökkenéshez vezet. Formálisan tehát minden rendszerinput automatikusan maximummal rendelkező nem monoton ceteris paribus alakzattal írandó le. A helyes kérdés tehát nem a hatásmechanizmus mikéntje, hanem az optimum-pontok attribútumonkénti helye, ill. az optimum-pontok ceteris paribus helyzetektől függő vándorlásának logikai helyessége.  Operatívan: Amennyiben iránykereső függvényillesztést alkalmazunk (MCM), akkor a lokális minimumok számának, ill. a polinomizálódás fokának (a hullám-természetű összefüggések számának) minimalizálása a cél. |

A statisztikai és logikai elvárások feltérképezése alapján:

* elvileg modellépítési kísérletekkel is ki kell tudni zárni mindennemű monoton alakzatot,
* vagyis csak lokális maximummal rendelkező átlagos rendszer-karakterisztikákat illene kapni ennek külön kikényszerítése nélkül is,
* így ennek hiányában a modellezés csak illegitim eredményre vezethet,
* következésképpen csak azok a társadalmi prevenciós stratégiák legitimek (vö. [B17]), ahol a valóság (vö. [B18], NaCl kb. 40%-a a Na, vagyis férfiak esetében kb. 7 mg/nap/fő, míg a nők esetében hazánkban 2009-ben mintegy 5 mg/nap/fő Na bevitel) és a feltárt optimumok viszonya egymással logikailag nem ütközik.

A következőkben előbb lássunk egy önmagában zárt regressziós vizsgálat sorozatot, majd egy hasonlóságelemzési bizonyítékláncot.

## Modellezési tapasztalatok

Az alábbiakban egy önálló, zárt elemzés következik a hagyományos matematikai-statisztikai eszköztár önkényesen/logikusan választott elemei és kapcsolataik alapján:

### Magyarázó modell többszörös lineáris regresszió ill. másodfokú polinomiális regresszió felhasználásával

#### Adatbázis

Független változók: X1 – X15 (tápanyag bevitel). Függő változó: X16 (várható élettartam) – vö. [B2]. Megfigyelési egységek: az országok adatai az EU15 és EU25 aggregátumokkal együtt 17 ország 5-5 időszaka esetében (85 db). A számítások az R (vö. [STAT2]) fejlesztői környezet segítségével kerültek elvégzésre.

#### A regressziós vizsgálatok célja

* Kimutatható-e összefüggés a magyarázó változók és a független változó között?
* Lineáris modellel leírható-e a kapcsolat?
* A függő változó változékonyságának mekkora részét képes magyarázni a megkonstruált modell?
* A magyarázó változók között van-e szoros összefüggés?
* A magyarázó változók között melyek azok, amelyeknek a függő változóra gyakorolt hatása a legjelentősebb?
* A só-bevitel milyen kapcsolatban áll a várható élettartammal?
* Jelentős-e a só-bevitel hatása a várható élettartamra?
* A felépített modellek közül melyiket tartjuk a legjobbnak? Milyen indokok alapján?

#### Alapmodell (M1) - többszörös lineáris regressziós modell

Az alapmodell felállításakor az összes független változó bevonásra került a lineáris regresszióba. Az illesztett lineáris modell adatai:

Residual standard error: 157.9 on 69 degrees of freedom

Multiple R-squared: 0.8229, Adjusted R-squared: 0.7844

F-statistic: 21.37 on 15 and 69 DF, p-value: < 2.2e-16

A modell magyarázóereje megfelelő (82%), a modell szignifikáns (p-value). A reziduumokra vonatkozó feltételek teljesülnek (1. ábra).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Inter-cept | X1 | X2 | X3 | X4 | **X5** | X6 | X7 | X8 | X9 | X10 | X11 | X12 | X13 | X14 | X15 |
| 686,3 | -1,9 | 0,7 | 0,5 | -24,0 | 0,0 | -36,7 | -1,3 | 49,7 | 73,9 | 185,4 | -82,0 | 0,4 | -2,6 | 45,1 | -10,6 |

1. táblázat: Az alapmodell változóinak koefficiensei (Forrás: saját számítások)

A kapott eredmények (vö. 1. táblázat) alapján a megadott változókkal jól jellemezhető a megadott független változó. Az ellenőrzés során varianciaelemzésre került sor:

C:\Ruff Feri\SZERVER\pitlik\Salt101228\1_modell_1_diag.tiff

1. ábra Az alapmodell-állítás (M1) feltételeinek grafikus ellenőrzése (Forrás: saját ábrázolás)

A modell vizsgálata ANOVA segítségével (2. táblázat):

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Response: | X16 |  |  |  |  |
|  | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
| X2 | 1 | 4076824 | 4076824 | 163,58 | 0,000 |
| X3 | 1 | 1310522 | 1310522 | 52,58 | 0,000 |
| X4 | 1 | 229688 | 229688 | 9,22 | 0,003 |
| **X5** | 1 | 456698 | 456698 | 18,32 | 0,000 |
| X8 | 1 | 257490 | 257490 | 10,33 | 0,002 |
| X10 | 1 | 671624 | 671624 | 26,95 | 0,000 |
| X11 | 1 | 825935 | 825935 | 33,14 | 0,000 |
| Residuals | 69 | 1719681 | 24923 |  |  |

2. táblázat: ANOVA-értékek (Forrás: saját számítások)

Az egyes változók hatását vizsgálva (t-érték) 4 magyarázó változó hatása szignifikáns (5%-os szignifikancia szinten). Eszerint a magyarázó változók közül néhány felelős a függő változó változékonyságának magyarázatáért. Közrejátszhat ebben egyes változók és az Y tényleges függetlensége, másrészt pedig a független változók között meghúzódó összefüggések. Ha megvizsgáljuk a modell (magyarázó) változóinak összefüggéseit (kollinearitás) pl. a VIF mutatóval, a következő értékeket kapjuk (vö. 3. táblázat):

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X1 | X2 | X3 | X4 | **X5** | X6 | X7 | X8 | X9 | X10 | X11 | X12 | X13 | X14 | X15 |
| 31,8 | 8,8 | 2,4 | 1,8 | **222,9** | 2,6 | 27,1 | 654,8 | 3,8 | 230,2 | 12,8 | 27,8 | 78,8 | 2,0 | 31,8 |

1. táblázat: VIF értékek (alapmodell) – (Forrás: saját számítások)

Erős kapcsolatok fedezhetők fel a független változók között, melyek az egyes változók hatásának felismerését teszik nehézzé, vagyis a megadott koefficiensek értékét kérdőjelezik meg.

#### Második modell (M2) - többszörös lineáris regressziós modell

A második modellt „stepwise” módszer segítségével került kialakításra az alapmodellből. Az eljárás végén a következő eredmény állt elő (vö. 4. táblázat):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Coefficients | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 3945,9 | 351,9 | 11,21 | 0,000 |
| X3 | 0,7 | 0,1 | 6,09 | 0,000 |
| X10 | 195,1 | 21,1 | 9,25 | 0,000 |
| X14 | 102,6 | 15,3 | 6,72 | 0,000 |
| X4 | -23,0 | 6,7 | -3,42 | 0,001 |
| X2 | 0,2 | 0,1 | 2,48 | 0,015 |

1. táblázat: 2. modell kiértékelése (Forrás: saját számítások)

A kapott eredmény (a modellbe bekerült változók) száma már lényegesen kisebb, mint az előző modellben. Itt minden változó szignifikáns hatást gyakorol a függő változóra. A második modell magyarázóereje valamivel kisebb (79%), mint az alapmodell esetében, de a modell most is szignifikánsnak bizonyult.

Residual standard error: 158.2 on 79 degrees of freedom

Multiple R-squared: 0.7965, Adjusted R-squared: 0.7836

F-statistic: 61.83 on 5 and 79 DF, p-value: < 2.2e-16

A magyarázó változók közötti kapcsolatokat jellemző mutató (VIF) értékei pedig a következők (vö. 5. táblázat):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| X3 | X10 | X14 | X4 | X2 |
| 1,67 | 1,08 | 1,43 | 1,56 | 1,57 |

1. táblázat: VIF értékek (2. modell) – (Forrás: saját számítások)

Ezek az értékek már elfogadhatóak (<5), azaz a modellbe bevont változók között nincs számottevő kapcsolat, amely a koefficiensek értékét befolyásolná.

A reziduumokra vonatkozó feltételeknek megfelel ez a modell is (2. ábra).

C:\Ruff Feri\SZERVER\pitlik\Salt101228\2_modell_1_diag.tiff

1. ábra: Diagnistic plots (M2) – (Forrás: saját ábrázolás)

A linearitás vizsgálatára a „component+residuals plot” (Chatterjee, 2006) grafikon került kialakításra (3. ábra). Megállapítható, hogy mindegyik változó esetében elfogadható a lineáris kapcsolat, bár a lokális regressziós görbe (3. ábra zöld grafikon) mutat némi „egyenetlenséget” az X4 és az X2 változó esetében, ami jelzi, hogy nem biztos, hogy a legjobb modellt határoztuk meg. A teszt alapján azonban elfogadható az a feltételezés, hogy a magyarázó változók és a magyarázott változó közötti kapcsolat lineáris összefüggéssel leírható.

C:\Ruff Feri\SZERVER\pitlik\Salt101228\2_modell_comp+resid.tiff

1. ábra: Component+residual plots (M2) – (Forrás: saját ábrázolás)

##### További háttérvizsgálatok

###### A modell kiválasztása

A fenti vizsgálatok elvégzése után a második modell került kiválasztásra további elemzések elvégzéséhez. A döntésben (vö. választás a hasonló magyarázóerővel rendelkező modellek közül) az játszott fontos szerepet, hogy a kevesebb változót tartalmazó modell változói közötti kapcsolatok kicsik, így az áttételes hatások kiismerhetetlensége elkerülhető.

###### A változók rangsorolása

A következő lépésben azt vizsgáljuk, hogy, melyik változónak milyen erős a kapcsolata az X16 (azaz Y) változóval. Ennek feltárására az R statisztikai szoftver „relaimpo” [STAT1] csomagja alkalmasnak bizonyul, mely többféle metrikát tartalmaz. Ezek közül a két legújabb került a mostani elemzésbe, melyek a változó fontosságának meghatározásakor figyelembe veszik a megmagyarázott eltérésnégyzet nagysága mellett a modellbe belépő változók sorrendjét is. Ezáltal egy nagyobb számításigényű (vö. a lehetséges permutációk végigszámolása), de több szempontot figyelembe vevő mutatókat kapunk.

|  |  |  |
| --- | --- | --- |
|  | lmg | pmvd |
| X10 | 0,326 | 0,374 |
| X14 | 0,259 | 0,355 |
| X2 | 0,179 | 0,061 |
| X3 | 0,195 | 0,176 |
| X4 | 0,041 | 0,034 |

1. táblázat: A változók jelentősége (M2) – Forrás: saját számítások

Mindkét mutató ugyanazt a sorrendet adja: legjelentősebb az X10 változó hatása, ezt követi az X14, … Az arányokban azonban eltérés mutatkozik (az értékek összege mindkét mutató esetében 1, tehát %-os formában értelmezhetők).

###### A kapott modell értelmezése

A változók ill. a változókhoz tartozó koefficiensek (a változók jelentőségének csökkenő sorrendjében)

|  |  |  |
| --- | --- | --- |
| X10 | Share of calories from protein (total) in total Dietary Energy Supply | 195,12 |
| X14 | Percentage shares of calories from poly-unsaturated fats | 102,57 |
| X3 | Fruit and vegetable consumption (grams/person/day) | 0,67 |
| X2 | Dietary energy supply (DES) (Kcal/p/d) | 0,20 |
| X4 | Non-Starch Polysaccharides (grams/person/day) | -22,98 |

1. táblázat: Változók sorrendje (Forrás: saját számítások)

* A legjelentősebb változó – a fehérjéből származó kalória aránya – kapcsolata pozitív az életkorral, és egy egységnyi növekedése (1%) a várható élettartamban 1,9 év növekedést jelent (ceteris paribus). A változó értékeinek terjedelme [10,6; 14,3]. Ebben a tartományban a kalóriarészesedés megfelelő, sőt még magasabb arány is kedvező élettani hatásokkal járna (vö. [STAT4]). Elfogadható tehát a növekvő lineáris kapcsolat ebben az intervallumban. A modellben nem maradt bent a másik két tápanyagforrásból (a zsírokból és a szénhidrátokból) származó energiarészesedés aránya. Megvizsgálva a három változó közötti kapcsolatot (korrelációs együttható) az látható, hogy a zsírokból (X9) ill. a szénhidrátokból (X8) származó energia-bevitel aránya egymással erős, negatív kapcsolatban van (r=-0,88), vagyis ezek helyettesítő viszonyban állnak egymással. A modellben megtalálható fehérjéből származó kalória aránya a szénhidrát részesedésével mutat negatív irányú kapcsolatot (bár itt r=-0,34), mely arra ad magyarázatot, hogy a 3 változó közül csak egy került be a modellbe. A WHO/FAO ajánlásokkal (vö. [STAT4]) szemben az X8 változó (szénhidrát) terjedelme 10 egységgel (%) „lejjebb” helyezkedik el, míg az X9 változó (zsírok) esetében a terjedelem 5 egységgel lefele és 10 egységgel fölfele is eltér.
* A második legjelentősebb hatással bíró változó (X14) a többszörösen telítetlen zsírsavakból származó energia aránya. Ez a változó a másik két – egyszeresen telítetlen és telített – zsírsavakkal kapcsolatos változókkal gyenge (r=0,3 ill. 0,12), de pozitív kapcsolatban van. A vizsgált változó (X14) a következő változókkal mutat erősebb kapcsolatot: - szénhidrátokból származó kalória aránya (negatív kapcsolat), - zsírokból származó kalória aránya (pozitív kapcsolat),- (pozitív kapcsolat). Ez utóbbi összefüggés azt jelzi, hogy a telítetlen zsírsavak növekedése esetén az zsírsav nagyobb ütemben növekszik, mint az zsírsav, így a kedvezőtlen arány (vö. [STAT5]) még inkább romlik.
* Jelentőségét tekintve a harmadik helyen a zöldség és gyümölcsfogyasztás áll. Ez a változó kapcsolatot mutat a következőkkel: egyszeresen telítetlen zsírsavakból származó energia (X6) – pozitív kapcsolat (p=0,68), cukorból származó kalória aránya (X7) – negatív kapcsolat (p=0,64). Az első kapcsolat magyarázata, hogy a telítetlen zsírsavak forrásai a növényi eredetű táplálékok (főként zöldség és gyümölcs). Az egyszeresen telítetlen zsírsavak főként növényi olajokban (olíva, napraforgó) és néhány gyümölcsben (mogyoró, mandula, dió, avokádó) találhatók meg. A cukorfogyasztással fennálló negatív kapcsolat a táblázatból is jól látszik. A mediterrán országok járnak az élen a gyümölcs és zöldségfogyasztásban, míg a cukorból származó energia-bevitel aránya a javasolt 10%-os érték alatt van (igaz, ez az évek során egyre közeledett a nem mediterrán területek értékéhez, ami a legutóbbi adatok szerint is fölötte van a határértéknek).
* Az étkezési energiaellátással is pozitív kapcsolatot mutat a várható élettartam, mely változó súlyát tekintve a negyedik helyre került. Az X2 változó pozitív irányú kapcsolatot mutat szénhidrát bevitellel, valamint a szénhidrátokból, a zsírokból, ill. a telített zsírsavakból származó energia arányával.
* A nem keményítő poliszacharid bevitelét mutató változó jelentősége bizonyult a legkisebbnek. Az eredmény érdekessége, hogy a változó, mely valószínűleg a cellulózt jelenti, negatív kapcsolatban áll az életkorral. A cellulózt az ember szervezete nem bontja, kiürül a szervezetből. Várakozásunk tehát az lenne, hogy nincs hatása a várható élettartamra. Igaz, jelentőségét tekintve az utolsó helyen áll (4,5%). A többi változóval való kapcsolatát vizsgálva 0,5 feletti korrelációs együtthatót csak a szénhidrát bevitellel és a gyümölcs- ill. zöldségbevitellel mutat. Ekkor viszont ellentmondásba kerülünk azzal, hogy a függő változó a gyümölcs és zöldség mennyiségével pozitív kapcsolatban van.

###### Lineáris részeredmények összegzése

A fenti összefüggések egy esettől eltekintve magyarázhatók, ill. az eddigi tudományos eredményekkel összhangban állnak. A problémás változó jelentősége kicsi, de akár további vizsgálatok végezhetők, melyben ennek a változónak kiküszöbölésével, esetleg új változó bevonásával vizsgálható, hogy ellentmondásmentessé tehető-e a modell. A magyarázó modellt ennek ellenére elfogadhatónak tarthatjuk, hiszen a függő változó változékonyságának nagy részét képes magyarázni a modell úgy, hogy a kapott eredmények eddigi ismereteinkkel összhangban állnak. A só-bevitelre vonatkozóan formálisan megállapítható, hogy:

* az M1 modell szerint: minél több a bevitt só, annál nagyobb a várható élettartam…
* az M2 modell szerint a só-bevitel nem gyakorol szignifikáns hatást a várható életkorra…

**A két modell egymással és látszólag az érvényes állami prevenciós politikával is ellentmondásban áll.**

#### Harmadik modell (M3) - polinomiális regressziós (másodfokú) modell

A változók lehetséges (elvárt) hatásait figyelembe véve polinomiális regressziót alkalmaztam, melyben minden változó második hatványát is belevettem a modellbe, majd stepwise módszerrel kialakítottam azt a modellt, melyben már minden változó hatása szignifikáns. Így az esetleges optimumok meghatározására nyílik lehetőség. A kapott modell a következő lett (8. táblázat):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| X2 | -4,16E+00 | 1,29 | -3,23 | 0,002 |
| X2^2 | 7,22E-04 | 0,00 | 3,68 | 0,000 |
| X3 | 6,75E-01 | 0,16 | 4,10 | 0,000 |
| X4 | 1,89E+02 | 58,75 | 3,22 | 0,002 |
| X4^2 | -4,06E+00 | 1,16 | -3,51 | 0,001 |
| **X5** | 4,90E-01 | 0,10 | 4,70 | 0,000 |
| **X5^2** | -7,83E-05 | 0,00 | -4,45 | 0,000 |
| X6^2 | -2,31E+00 | 0,62 | -3,71 | 0,000 |
| X8^2 | 3,61E-01 | 0,12 | 2,95 | 0,004 |
| X9 | 3,14E+02 | 47,72 | 6,57 | 0,000 |
| X9^2 | -3,34E+00 | 0,71 | -4,69 | 0,000 |
| X10^2 | 4,45E+01 | 19,84 | 2,24 | 0,028 |
| X11 | -9,33E+01 | 14,45 | -6,46 | 0,000 |
| X13^2 | -1,19E-02 | 0,005 | -2,27 | 0,026 |

8. táblázat: Modell-paraméterek (M3) – (Forrás: saját számítások)

|  |
| --- |
| Residual standard error: 116.5 on 69 degrees of freedom |
| Multiple R-squared: 0.9035, Adjusted R-squared: 0.8826 |
| F-statistic: 43.09 on 15 and 69 DF, p-value: < 2.2e-16 |

Az eredményből látszik, hogy a négyzetes tagok szignifikánsak (5%-os szinten), együtthatójuk negatív, ill. pozitív volta jelzi, hogy a függvényeknek lokális maximuma, ill. minimuma van. A modell szignifikánsnak bizonyult, magyarázóereje nagy. A reziduumokra vonatkozó szükséges feltételek teljesülése a 4. ábrán látható.

A változók közötti kapcsolatok erősek (mérés: VIF), hiszen több változó esetében annak négyzetét is tartalmazza a modell. Emiatt a lineáris modellnél alkalmazott, a változók fontossági sorrendjét meghatározó, algoritmus nem vezetett eredményre.

C:\Ruff Feri\SZERVER\pitlik\Salt101228\3_modell_1_diag.tiff

1. ábra: Diagnostic plots (M3) – (Forrás: saját ábrázolás)

##### További vizsgálatok

###### A modell értelmezése

Amely változó esetében a lineáris tényező kiesett a modellből (X6, X8, X10) ott matematikailag a parabola csúcspontja a változó 0 értékéhez tartozik, vagyis a nemnegativitási feltételt figyelembe véve a függvény az adott tartományban monoton növekvő (X8, X10) ill. monoton csökkenő (X6), de nem lineáris a kapcsolat, hanem másodfokú. Ezek az eredmények a lineáris alapmodellel egyezést mutatnak a növekedés, ill. a csökkenés tekintetében.

A többi másodfokú kifejezés esetén a szélsőérték helyek a következők: X2: 2882,7 - X4 : 23,3 - **X5 (só-bevitel) : 3126,8 (mg/p/d/1000, ahol a vizsgált minta minimuma 1315 mg/p/d/1000, ill. maximuma 5404 mg/p/d/1000, az optimum felett 10 objektum található a 85-ből: 1963PT, 1973PT, 1993DK, 1993PT, 2003DK, 2003FI, 2003GR, 2003NL, 2003PT, 2003SE)** - X9 : 47. Ezek közül az X2 változó esetében lokális minimumról, a többi esetében pedig lokális maximumról beszélünk (ld. a négyzetes tagok koefficiense).

###### A lokális maximumok értelmezése

X4 – nem keményítőből származó poliszacharid bevitel, X5 – só fogyasztás, X9 – zsírokból származó energia bevitel. Itt a bevitt mennyiség és az élettartam közötti másodfokú kapcsolat optimális mennyiséget mutat. Azonban az X9 változónál a maximumhely kívül esik a vizsgált tartományon (20,3-41,5), így monoton növekvő függvény szerint határozza meg a függő változó értékeit. Összevetve a lineáris modellel azt tapasztaljuk, hogy az X4-es és X9-es változóval negatív korrelációt mutatott a függő változó, míg a só-fogyasztás nem is szerepelt benne. Másodfokú összefüggést feltételezve viszont az X5 a modell egyik szignifikáns változójává vált. Az optimális beviteli értékek létezése az adott változók esetében elfogadható, az eddigi tapasztalatokkal egyezik. A felvetődő kérdés inkább az, hogy a többi változó esetében: *Miért nem kaptunk ilyen optimális értéket?* pl.

* A független változó vizsgált tartományában még nincs benne az optimális függvényértéket adó érték (ill. annak egy környezete).
* A független változó növekedése az anyagi jólétet képezi le, és más változók (életvitel, egészségügyi ellátások, táplálék kiegészítők) figyelembe nem vett hatása ezt kompenzálja a várható élettartam tekintetében.

###### A lokális minimumok értelmezése

X2 – energia-bevitel: A lokális minimumok magyarázata nehéz. *Lenne olyan szintje a független változónak, ahová minimális várható élettartam tartozik, és attól balra is és jobbra is nagyobb a függvényérték?* Ennek elfogadása az eddigi ismereteinkkel ütközne, hacsak a 85 objektum nem bomlik a lokális minimum körül 2 logikai szinten felismerhető részhalmazra (pl. térbeli, kulturális, időszakos, ill. egyéb jelleg alapján). Ennek vizsgálata azonban már meghaladta a tervezett tanulmány kereteit…

#### Negyedik modell (M4) - polinomiális regressziós (másodfokú) modell

Mivel a harmadik modellben a változók közötti kapcsolatok erősek voltak, a következő modellben ennek kiküszöbölését célozzuk meg. A független változóknak egy lineáris transzformációját hajtom végre: a változók értékeiből kivontuk azok átlagát.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| X2^2 | 1,05E-03 | 0,000 | 4,838 | 8.02e-06 |
| X3 | 7,76E-01 | 0,121 | 6,429 | 0,000 |
| X4^2 | -4,31E+00 | 1,101 | -3,911 | 0,000 |
| **X5** | 1,98E-01 | 0,040 | 5,01 | 0,000 |
| **X5^2** | -1,26E-04 | 0,000 | -6,366 | 0,000 |
| X6^2 | -8,15E+00 | 2,633 | -3,096 | 0,003 |
| X8^2 | 3,94E+00 | 1,402 | 2,812 | 0,006 |
| X9 | 1,94E+01 | 7,465 | 2,594 | 0,012 |
| X9^2 | -5,60E+00 | 1,369 | -4,092 | 0,000 |
| X10 | 1,04E+02 | 26,530 | 3,912 | 0,000 |
| X10^2 | 7,33E+01 | 18,180 | 4,034 | 0,000 |
| X11 | -7,22E+01 | 14,530 | -4,971 | 0,000 |
| X11^2 | 9,39E+00 | 3,648 | 2,574 | 0,012 |
| X12 | 1,28E+00 | 0,398 | 3,217 | 0,002 |
| X12^2 | -9,12E-03 | 0,002 | -3,767 | 0,000 |
| X14^2 | 2,04E+01 | 5,859 | 3,487 | 0,001 |

9. táblázat: Modell-paraméterek (M4) – (Forrás: saját számítások)

Residual standard error: 116,2 on 67 degrees of freedom

Multiple R-squared: 0,9068, Adjusted R-squared: 0,8832

F-statistic: 38,36 on 17 and 67 DF, p-value: < 2,2e-16

Modell felállításának szükséges feltételeit a Diagnostic plot alapján teljesítettnek fogadtuk el.

##### További vizsgálatok

###### A lokális maximumok értelmezése

A modell esetében négy változónál találunk lokális szélsőértéket a megfigyelési tartomány belsejében: X5, X9, X11, X12. Ezek közül a harmadik minimumhelyet ad, a többi maximumhelyet. A szélsőérték helyek a következőképpen alakultak:

**X5: 789,18+2360,09=3149,27 (mg/p/d/1000)** - X9: 1,73 + 34,84 = 36,57 - X11: 3,85 + 12,03 = 15,88 - X12: 70,20 + 403,81 = 474,01.

A lokális maximumok értelmezésekor az első különbséget abban látjuk (M3-hoz képest), hogy az X9 változó esetében a maximumhely már a vizsgált tartományba esik, vagyis valóban van optimális értéke. A só hatásának megítélése ugyanaz, mint az M3-ban. Az X12 változó pedig nem szerepelt az előző modellben. . Az optimális beviteli értékek létezése ismételten az eddigi tapasztalatainkat, elvárásainkat testesíti meg.

###### A lokális minimumok értelmezése

A lokális minimum kérdése ismét felmerül az X11 (koleszterinből származó kalória) változó kapcsán. Azonban itt a szélsőérték hely vizsgálatából az derül ki, hogy annak jobb oldalán csak 2 db megfigyelési egység található (1963FI, 1963UK), amely azonban nem meggyőző e tekintetben. Így az X11 vizsgált tartományában inkább monoton csökkenő függvénnyel jellemezhető a kapcsolat.

###### A változók sorrendje

A változók sorrendjének meghatározásakor ismét a lineáris modellnél bemutatott értékeket használtuk (10. táblázat):

|  |  |  |
| --- | --- | --- |
|  | lmg | pmvd |
| X3 | 0,155 | 0,145 |
| X10 | 0,140 | 0,210 |
| X9 | 0,130 | 0,229 |
| X12 | 0,085 | 0,051 |
| **X5** | 0,077 | 0,055 |
| X11 | 0,072 | 0,079 |

10. táblázat: Változók sorrendjének megállapítása (M4) – (Forrás: saját számítások)

A modell meghatározó változói: X3, X10, X9. A két mutató más sorrendet ad, de ez a három a legjelentősebb mindkettő szerint. Ezek közül az X10 és az X3 a lineáris modellek esetében is benne van az első három között, a változók fontosságát tekintve.

###### Genetikai potenciál számítása

Minden egyes modellbeli változó esetén megkerestük a (maximális) függvényértéket adó elemet az adott változó értelmezési tartományában - ez lineáris becslés esetén az intervallum széle, másodfokú függvény esetén pedig vagy közbülső érték (maximum esetén), vagy az intervallum valamelyik széle (minimum esetén).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| M3 | coeff | *x*max | *x*min |  |
| X2 | -4,5290 | 3798 | 2972,8 |  |
| X2^2 | 0,0008 |  |  |  |
| X3 | 0,6931 | 1211 | 167 |  |
| X4 | 183,5000 | 23,3 | 33,3 |  |
| X4^2 | -3,9290 |  |  |  |
| **X5** | 0,4732 | 3126,8 | 5404 |  |
| **X5^2** | -0,0001 |  |  |  |
| X6^2 | -2,3360 | 9,6 | 20,5 |  |
| X8^2 | 0,3709 | 63,2 | 44,8 |  |
| X9 | 316,0000 | 41,5 | 20,3 |  |
| X9^2 | -3,3720 |  |  |  |
| X10^2 | 53,3200 | 14,3 | 10,6 |  |
| X11 | -90,3000 | 5,5 | 16,4 |  |
| X13^2 | -0,0116 | 191 | 269 |  |

11. táblázat: Genetikai potenciál számítása (M3) - (Forrás: saját számítások)

Becslő függvény értéke: xmax-nál: y =10775 azaz 107,75 év. xmin-nél: y =2946 azaz 29,46 év.

|  |  |  |  |
| --- | --- | --- | --- |
| M4 | coeff | *x*max | *x*min |
| X2^2 | 1,05E-03 | 3798 | 2565 |
| X3 | 7,76E-01 | 1211 | 167 |
| X4^2 | -4,31E+00 | 18,3 | 33,3 |
| **X5** | 1,98E-01 | 3149,27 | 5404 |
| **X5^2** | -1,26E-04 |  |  |
| X6^2 | -8,15E+00 | 9,6 | 20,5 |
| X8^2 | 3,94E+00 | 63,2 | 44,8 |
| X9 | 1,94E+01 | 36,57 | 20,3 |
| X9^2 | -5,60E+00 |  |  |
| X10 | 1,04E+02 | 14,3 | 10,6 |
| X10^2 | 7,33E+01 |  |  |
| X11 | -7,22E+01 | 11,1 | 15,88 |
| X11^2 | 9,39E+00 |  |  |
| X12 | 1,28E+00 |  |  |
| X12^2 | -9,12E-03 | 474,01 | 155 |
| X14^2 | 2,04E+01 | 9,1 | 2,8 |

12. táblázat: Genetikai potenciál számítása (M4) - (Forrás: saját számítások)

Becslő függvény értéke: xmax-nál: y =10323 azaz 103,23 év. xmin-nél: y =3922 azaz 39,22 év.

#### A modellek számszerű összehasonlítása

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model | AIC | R2 | adj R2 | VIF | F test (5%) | Resid. st. err. | Pearson | Ymax | Ymin | Ytény=Ybecs | Fontosság |
| M1 | 1118 | 0.8229 | 0.7844 | bad | OK | 157.9 | 0.907 | 64.11 | 35,92 | OK | X10, X14, X3 |
| M2 | 1110 | 0.7965 | 0.7836 | good | OK | 158.2 | 0.892 | 88,38 | 61,7 | OK | X10, X14, X3 |
| M3 | 1066 | 0.9035 | 0.8826 | bad | OK | 116.5 | 0.950 | 107,75 | 29,46 | OK | - |
| M4 | 1067 | 0.9068 | 0.8832 | good | OK | 116.2 | 0.952 | 103,23 | 39,22 | OK | X3, X10, X9 |

13. táblázat: A modellek jellemzésére használt mutatók értékei (Forrás: saját számítások)

Az egyes változók hatása az Y-ra:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | M1 | M2 | M3 | M4 |
| X1 | - |  |  |  |
| X2 | + | + | min | + |
| X3 | + | + | + | + |
| X4 | - | - | max | - |
| **X5** | + |  | max | max |
| X6 | - |  | - | - |
| X7 | - |  |  |  |
| X8 | + |  | + | + |
| X9 | + |  | + | max |
| X10 | + | + | + | + |
| X11 | - |  | - | ?min? |
| X12 | + |  |  | max |
| X13 | - |  | - |  |
| X14 | + | + |  | + |
| X15 | - |  |  |  |

14. táblázat: Ceteris paribus paraméterek (Forrás: saját számítások)

#### Következtetések

A legjobb modell: M4. A modell a paraméterek összességét megfigyelve, lényegében minden tekintetben a legjobb értékeket veszi fel. A só ceteris paribus alakzata: egy olyan másodfokú görbe, mely esetében az értelmezési tartomány egy belső pontjához (3149 mg/p/d/1000) tartozik a függvény maximális értéke (csúcspont). A jelenlegi prevenciós kampány legitimitása: A kapott modell kvázi ellentmondásmentes, ami a modell elfogadásának alapfeltétele, hacsak az új hatásmechanizmusok kísérletekkel utólag alá nem támaszthatók. A modellek az egyes változók hatását egyformának találták, kivéve természetesen az optimumok esetét, melyeket a lineáris modellek nem tudnak felismerni. **A sóra vonatkozó optimum-pont és a magyar valóság alapján a jelenleg érvényes állami prevenciós stratégia matematikai-statisztikai szempontból legitimnek nevezhető. Bár az M4-es modell is fennakad formálisan a lokális minimumok rendszerelméleti KO-feltételén.**

### Hasonlóságelemzés

A hasonlóságelemzés lényege - a szakirodalmi rövid leírás alapján - a lépcsős függvények alkalmazása (vö. [B19]). Egy lépcsős függvény felhasználása feltételezi, hogy a tetszőlegesen skálán ábrázolt inputok (független változók) rangsorskálára kerülnek letranszformálásra, ahol a sorszámozás iránya egyben a ceteris paribus alakzatok előjelének megfelelője. Ha feltételezhető, hogy egy X és az Y között a kapcsolat a ’minél nagyobb annál nagyobb elvnek’ felel meg, akkor az az inputadat kapja ezen változó esetén az egyes rangsorszámot (a legnagyobb felhajtó erő jelét), amelyik a legnagyobb az összes megfigyelés között. A ’minél kisebb annál nagyobb elv’ esetén természetesen a legkisebb inputérték lesz az egyes rangsorszámot viselő adott változó esetén. Abban az esetben, ha az irányok nem adhatók meg előre (vö. tanulmány célja: iránykereső modellezés), akkor a lépcsők monotonitási feltétele, vagyis ’a jobb rangsorszámhoz nem tartozhat kisebb lépcsőérték (felhajtó erő), mint egy ennél gyengébb rangsorpozícióhoz elv’ feloldásra kerül. A „lépcsők” helyén így formálisan tetszőleges alakzatok (vö. polinomok, nem szimmetrikus görbék) keletkezhetnek. A lépcsős függvények ceteris paribus alakzatai alapértelmezés szerint tehát rendszerszintűek, azaz egy változó adott szintjének hatása a többi keretfeltételtől független konstans. (Természetesen kellően sok inputmegfigyelés esetén a tanulási minta szegmentálása tetszőleges logikák mentén megoldható, vagyis az eredetileg egyetlen konstansként reprezentált átlagos hatásmérték és irány részekre bontható).

A lépcsők megfelelnek a döntési fák egy csoportba sorolt objektumainak (vö. node), de a hasonlóságelemzés esetén előre dönteni kell a lépcsők számáról és ezek formális küszöbértékeiről. A lépcsőfokok össevonása lehetővé teszi a lépcsőszám redukciót, ami alapértelmezés szerinti futtatások esetén (vö. minden objektum egy lépcsőfok), az ideális lépcsőszám és küszöbérték-rendszer optimalizálását jelenti.

A becslési érték adott megfigyelés (objektum) primer inputadatai helyére lépő rangsorszámoknak megfelelő lépcsőfokok összegeként értelmezhető (vö. additív modell). Más megfogalmazásban a lépcsős függvény egy szakértői rendszer, mely inputja tetszőleges számú változó tetszőleges számú opciója (lépcsőfoka). Az ezekből előálló tetszőleges variációhoz (vö. kombinatorikai tér) a kimeneti érték mindenkor a lépcsőszintek helyettesítési értékeinek (a felhajtóerőknek) az összege.

A lépcsőfokokat egy LP (lineáris programozási) modell számítja ki, melyben (abszolút értékek és hibanégyzet számítások kizárása esetén) a célfüggvény kettős: egyrészt a becslések és a tények, ill. tények és becslések pozitív és negatív távolsága külön-külön futtatásban, de egyaránt legyen minimális. A korlátozó feltételek (ha vannak), akkor csak annyit írnak elő, hogy a lépcsők legyenek lépcsők, vagyis a jobb rangsorszám helyettesítési értéke nem lehet kisebb, mint egy rosszabb rangsorszám esetén. Így természetesen lépcsőfokok összevonódása alakulhat ki, ha a tények és a becslések távolságának minimalizálása ezt kívánja meg. A hasonlóságelemzés lépcsős függvényeinek paramétereit (konstansait) egy LP modell vezeti le közelítő jelleggel. A hasonlóságelemzés tehát alapértelmezés szerint optimalizált kombinatorikai problémával, diszkrét becslési fennsíkokká alakítja át az esetlegesen folytonos problémateret.

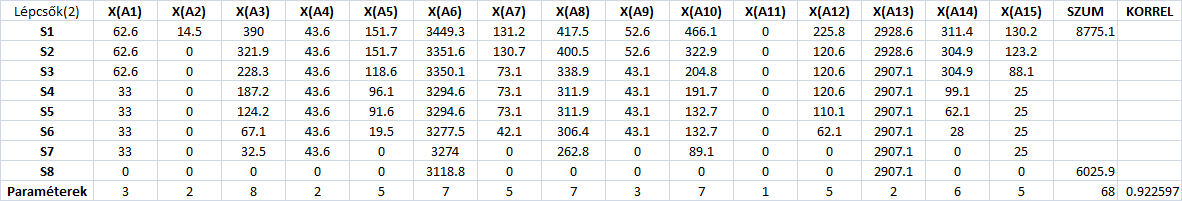
Az additív modelleken túl léteznek multiplikatív modellek, melyekben a lépcsőfokok szorzata vezet a becslési értékhez. S természetesen tetszőleges hibrid-modellek is léteznek, melyekben tetszőleges összevonási eljárás állhat a becslés mögött, ill. maguk a primer adatok is hatással lehetnek részei a becslési függvénynek. Jelen tanulmány vizsgálatai additív és multiplikatív jellegűek voltak.

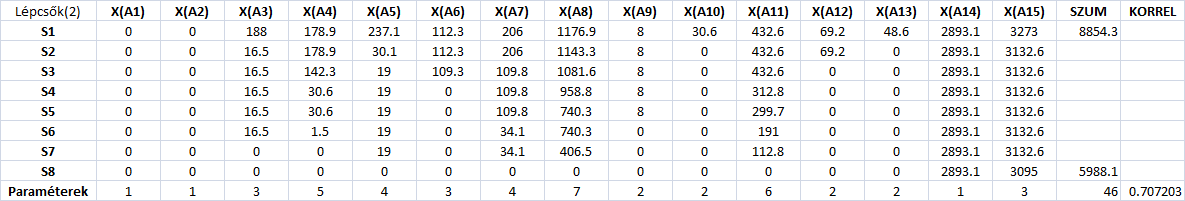
A hasonlóságelemzés képes a legkisebb négyzetek elve mentén is a célfüggvény szélsőértékének keresésére. Ez azonban speciális jelentéstartalommal bír az abszolút hiba-minimum, vagy az előjeles hibák kétrétegű minimalizálásával szemben: a négyzetes hiba esetén a nagyobb hibák egy egységgel való csökkentése jobban honorálódik, mint az alacsonyabb szintű hibák egy egységgel való csökkentése. Vagyis azt sugallja a hiba definíciója, hogy a független változók felelnek elsődlegesen és alapvetően a függőváltozó változásaiért….

#### Monoton additív modellek

##### Inverz modellek üzenete

Amennyiben arra a kérdésre keressük a választ, vajon bármilyen monoton összefüggés feltárható-e a függő és a független változók között, úgy a legegyszerűbb teszt egy tetszőleges előjel-sorozat (minden pozitív) és ennek inverze (minden negatív) esetén a modellek kialakítása. Elsőként a 85 objektum minden egyes sorszáma önálló paraméterhelyként került meghatározásra (15\*85 potenciális paraméterhellyel). Ebben az esetben a korreláció a tények és a becslések között: KORREL(direkt)= 0.9999987, KORREL(inverz)=0.944289539.





HE1.táblázat: Direkt és inverz azonos irányú modellek pontossága (Forrás: saját számítások)

A táblázati értékek mértékegysége: év/100

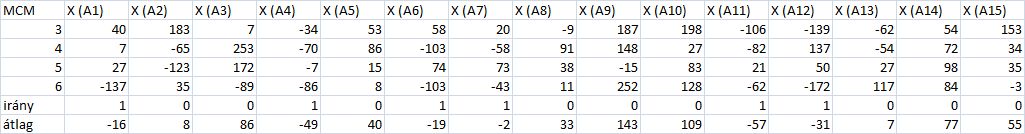
Amennyiben büntetni kívánjuk a nagy paraméterszámot, akkor a hasonlóságelemzés keretében tetszőleges szintre csökkenthető a „felbontás”, jelen esetben 8 lépcsőre került definiálásra az eredeti rangsorszámok (r1) alapján (r2= int(r1/11)+1). Jól látható (HE1. táblázat), hogy a korreláció az inverz modell esetében jelentősen esett, de ezzel együtt a valós paraméterszám is csökkent (68-ról 46-ra). A genetikai potenciál értékei (87.751 év, ill. 60.259 év, valamint 88.543 év, ill. 59.881 év) nem változtak jelentősen. Az inverzió hatására 85-ből 47 esetben megfelelő volt a tény-becslés különbségének irányváltás, 38 esetben azonban nem. Vagyis a jelenség (a modellezendő rendszer) mintegy 45%-ban vakfoltokat eredményezett a 85-lépcsős közelítésben. A 8 lépcsős felbontás mellett a vakfoltok száma eggyel meghaladta az éles, azaz tükrözött előjelű objektumok számát.

Az inverz modell korrelációja gyengébb, de ezt kevesebb paraméterrel éri el (vö. adjusted R^2), azaz látszólag közel azonos stabilitást lehet feltételezni ebben az esetben is. A direkt (minden változó esetén a minél nagyobb annál nagyobb elvet hasznosító) modell esetében minden - az X11 kivételével - változó bevonásra került a modellbe, ami a hasonlóságelemzés logikája szerint azt sugallja, hogy ez a modell-logika hitelesebb, hiszen a hasonlóságelemzésből kieső változók azért válnak zajjá, mert úgy, ahogy hatniuk szabad nem képesek hatni. Ezt erősíti az X11 viselkedése: mely az egyik modellből kiesik, míg a másik modellben magas abszolút értékekkel és nagy lépcső-szórással bír.

**Konklúziók: A vizsgált változók inkább hatnak pozitív előjel mentén a várható élettartamra, mint fordítva. A só-bevitel esetén az élettartam-növelő monoton hatás 8%-os fontossággal volt igaz a 15 mutatószám által kirajzolt 100%-on belül 85 lépcső esetén, míg az inverz (csökkentő) hatás 2%-os fontosságot mutatott csak (85 lépcső esetén, ill. 8-8 lépcső esetén 1-1% volt ez a hatásmérték). Az állami stratégia ezek alapján illegitim.**

##### Másodlagos monoton irányok

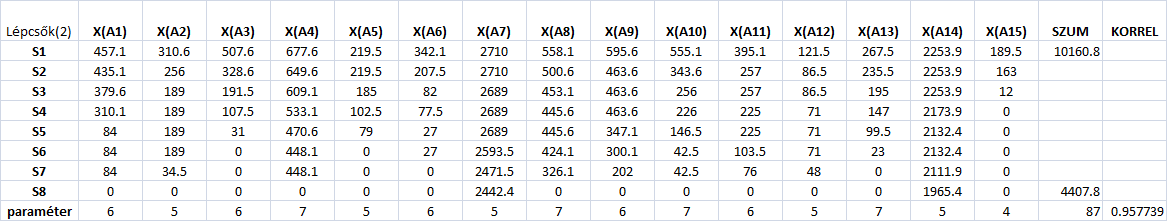
Amennyiben iránykereső eljárások segítségével a legvalószínűbb másodlagos irányt határozzuk meg, akkor az alábbi irányok ismerhetők fel (HE2. táblázat):

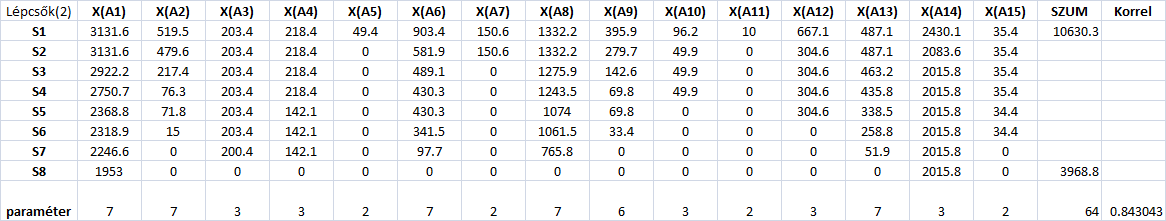


HE2. táblázat: Másodlagos iránykeresés eredményei (Forrás: saját számítások)

A táblázati értékek mértékegysége: év/100

A vegyes irányú modell és inverze szintén felépítésre került 85 és 8 lépcsővel egyaránt (vö. HE3. táblázat):

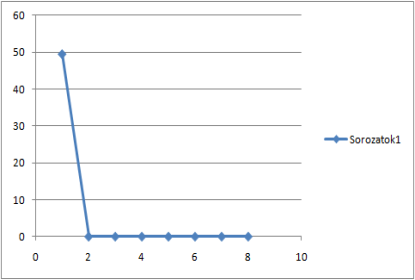
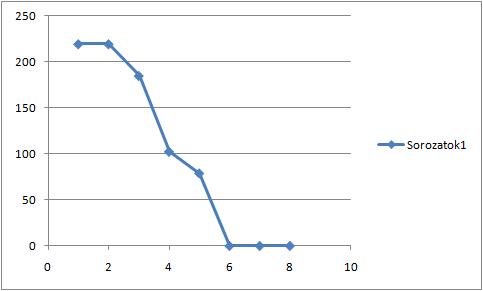




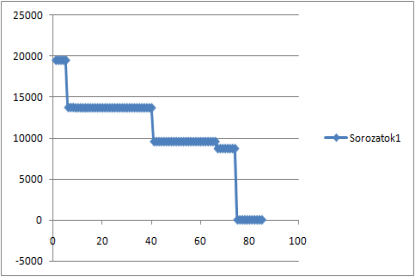
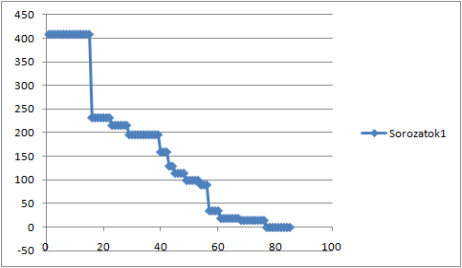
HE3.táblázat: Direkt és inverz vegyes irányú modellek pontossága (Forrás: saját számítások)

A táblázati értékek mértékegysége: év/100

A vegyes irányok hatására a modellek stabilitása jelentősen nőtt. Az inverzió éles és vak aránya 43:42-ről 85 lépcső esetén 41:44-re romlott 8 lépcsőt alkalmazva. A só fontossága a direkt vegyes irányú modellben 85 lépcső esetén már csak 3% volt (8 lépcső esetén 1%), míg az indirekt vegyes 85 lépcsős modellben 2% (a 8 lépcsős modellben 0%). Vagyis a fontosság stabilitása a direkt modellben a többi változó és ezek módosított iránya esetén nem adott az azonos irányú modellekhez képest növekvő korrelációk mellett: KORREL(direkt\_85)=0.9999, KORREL(indirekt\_85)=0.99999.



Modell: direkt vegyes 8 lépcsős // indirekt vegyes 8 lépcsős



Modell: direkt vegyes 85 lépcsős // indirekt vegyes 85 lépcsős

HE1. ábra: Só (X, lépcsőszint) és élettartam (Y, év/100) összefüggések lépcsőszám és inverzió (Forrás: saját ábrázolás)

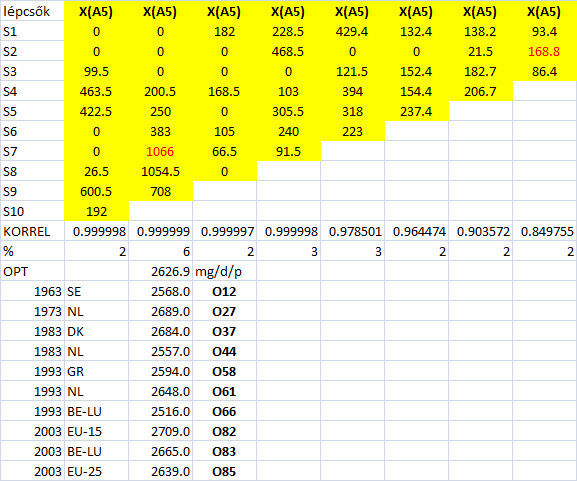
Konklúziók: A vegyes irányú modellek esetén a sóra és az élettartamra vonatkozó egyenes arányossági hatásmechanizmus valószínűbb, de ennek stabilitása nem egyértelmű a véletlenszerűen választott azonos irányú modellekhez képest, hiszen

* a magasabb korrelációs értékekhez (+)
* több paraméterhelyre volt szükség (-),
* a genetikai potenciál szélesebb intervallumot írt le (+),
* **a só aktív lépcsőinek száma az indirekt esetben kevesebb, szinte zajt jelentő (+),**
* a só fontossága a vegyes direkt modellben kisebb és instabil (-),
* a 85 lépcsős modellben az inverz alak volt az erősebb (-).

Mindösszesen az **iránykeresés a só esetén minden esetben az egyenes arányosságot mutatta ki**, ill. az inverz vegyes 8-lépcsős modellben a só szinte zajjá vált: vagyis a só inkább egyenes, mint fordított arányosságot mutat, ha monoton tendenciát kellene valószínűsíteni, de ezen értékítélet stabilitás a modellezés túltanulás-orientált alaplogikája szerint inkább egy fel nem vállalt szakértői nyilatkozat jelent, mint sem elkötelezett kiállást a valószínűbb hatásmechanizmus mellett.

#### Nem monoton additív modellek

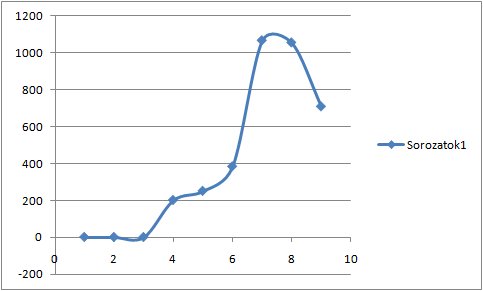
Elfogadva, hogy a monoton irányok feltárására törekvő modellezés végső szakértői véleményre nem vezet, iránykereső modellezésre kell átváltani. Egészen pontosan eleve iránykereső modellezést kell elsőként végrehajtani, melyet a monoton modellekkel lehet visszaellenőrizni. Itt és most csak a regressziós esettanulmány szerkezetének követése okán alakult ki ez a sorrend.



HE4. táblázat: Iránykereső modellezés eredményei (Forrás: saját számítások)

A felső táblázati értékek mértékegysége: év/100,

az alsó táblázatban a Na-beviteli adatok 1000-szerese látható (mg/d/p/1000)



HE2. ábra: Az egyedül értelmezhető/legitim ceteris paribus alakzat a só (X) és a várható élettartam (Y, év/100) esetében (Forrás: saját számítások)

A HE4. táblázat és a HE2. ábra alapján kijelenthető:

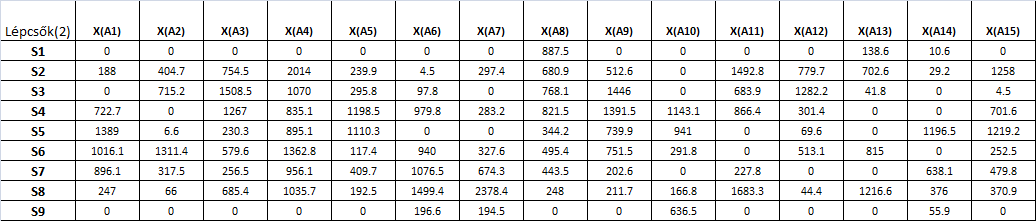
* a 3. és a 9. lépcsős modell kivételével jelentős a polinomizálódási modell szintű hajlam, melyet jelen esetben csak a só-bevitelre vonatkozó rétegek reprezentálnak a teljes lépcső-szerkezetek helyett
* a legjobb modell (matematikai értelemben véletlenül) éppen a klasszikus egy szélsőértékkel rendelkező modell,
* mely nem szimmetrikus optimumként ismeri el és fel a só hatását, vagyis a többlet-bevitel esetén a hatásmérték kevésbé drasztikus, mint hiány esetén,
* az optimum mértéke: 2626.9 mg/d/p/1000 a kapcsolódó objektumok (S7, 1066-os helyettesítési érték) alapján.
* A só fontossága a 9 lépcsős MCM modellben 6%, míg minden egyéb esetben 2-3% között ingadozik.

Konklúziók:

* a modelleredményekből kikényszeríthető a keresett válasz,
* de ennek stabilitása alacsony szintű,
* vagyis a só és a várható élettartam kölcsönhatása optimum jellegű,
* s a magyar só-beviteli adatok magasan az optimum feletti tartományba esnek,
* **vagyis az állami prevenciós stratégia legitim egy instabil minősítés eredményeként.**

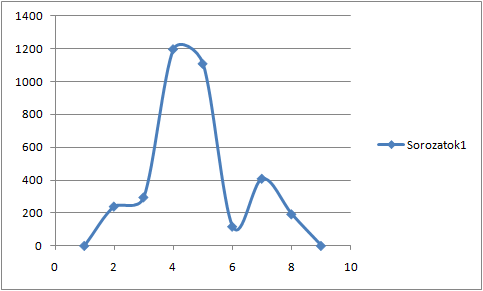
#### Szekunder adatos additív modellek

A szekunder adatok bevonása a modellezés stressz-tesztje, amikor is elvileg ismert, gyakorlatilag nem mért extrém input-hiányos és input-túlterheléses élethelyzetek következményeit kell a primer megfigyelésekkel együtt kezelni tudni.



HE5. táblázat: 87 objektumos additív MCM modell eredményei (Forrás: saját számítások)

Mértékegység: év/100



HE3. ábra: A só (X) hatásmechanizmusa (Y=várható élettartam, év/100) szekunder adatok alapján (Forrás: saját számítások)

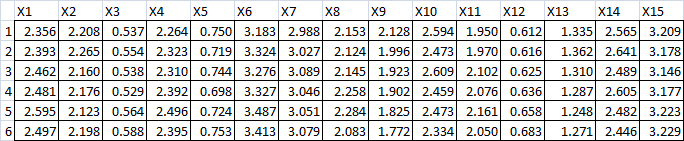
A só esetében a polinomizálódás foka és mértéke alacsony (vö. HE3. ábra). A többi változó esetében általános érvényű optimum-hatásról nem lehet beszélni, de még a monotonitás sem tipikus. A modell korrelációs értéke: 0.951557263 A genetikai potenciálok 214 és 0 év között ingadoznak. A só fontossága: 6%. Az optimum szintje a bevitel mediánjához közelít (2240 mg/d/p/1000), s nem szimmetrikus, vagyis többletbevitel esetén a hatásmérték kevésbé drasztikus mint hiány esetén.

**Konklúzió: a magyar prevenciós politika instabil érvrendszer mellett, de inkább legitim, mint sem.**

#### Multiplikatív modellek

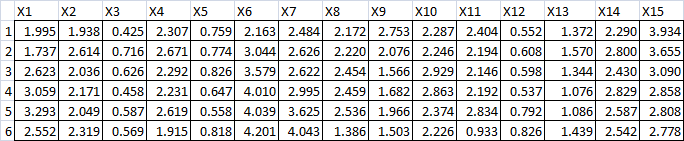
A multiplikatív modellek bevezetésének oka, hogy az élettani folyamatok (bár közelíthetők additív módon) alapvetően multiplikatív jellegűek, vagyis már milyen tényező hiánya a rendszer összeomlásához kell, hogy vezessen. (A 6 lépcsős modell jelenleg a futtathatóság maximumát jelenti 15 mutatószám mellett Excel solver-támogatással).

A 85 soros primer megfigyeléseket feldolgozó modell nem került szembesítésre azzal a rendszerelméleti többlettudással (+2 rekord: végig 1 és végig 6 sorszámokat tartalmazó rekordok), mely szerint bármilyen input eltúlzása vagy hiánya nulla várható élettartamot von maga után.



HE6. táblázat: 85 objektumos 6 lépcsős multiplikatív MCM modell lépcsői

(Forrás: saját számítások)



HE7. táblázat: 87 objektumos 6 lépcsős multiplikatív MCM modell lépcsői

(Forrás: saját számítások)

A lépcsők megfelelő szintjeinek szorzata a várható életkor százszorosát adja.



HE8. táblázat: Modell-adatok (Forrás: saját számítások)

A genetikai potenciál adatok ’év’-ben értendők.

A HE6-8. táblázatokból levonható értelmezések:

* az X3, az X5 és az X12 változók semelyik értéke nem éri el az 1-értéket, vagyis ezen változók semmilyen szintje nem növeli a várható élettartamot, csak kisebb nagyobb mértékben csökkenti a többi tényező által elérhetővé tett potenciált, ami önmagában a só-csökkentés szinonimája,
* a X5 hatásmechanizmusai polinomizálódóak, vagyis irracionálisak,
* a 87-elemű tanulási minta lényegében értelmezhetetlen a rendszer számára a szélsőséges genetikai potenciálok miatt,
* a multiplikatív modellek lépcsői nem tükröznek minden attribútumra optimum-jellegű lefutásokat, lényegében még a monotonitás is esetleges.

Konklúziók:

* A só és a várható élettartam között immár egy instabil fordított arányosság is feltételezhető,
* **ami az állami prevenciós stratégia igazát támasztja alá instabil bizonyítási folyamat eredményeként.**

#### A hasonlóságelemzési modellek együttes értékelése

**Instabil bizonyítási kísérletek alapján az állami prevenciós stratégia legitimitása mellett több érv szól, mint ellene. A bizonyítási kísérletek ellentmondásai:**

* **a monoton modell egyenes arányosságot valószínűsít,**
* **az additív nem monoton modell optimum-hatást sejtet,**
* **a multiplikatív modell a só bevitel hatását minden mennyiségben negatív hatásúnak mutatja,**
* **a só alapvetően nem tartozik a fontos modell-változók közé,**
* **az inverziós vakfoltok aránya magas,**
* **az optimum-hatások aszimmetriája a prevenciós stratégia ellenében hat,**
* **az egymásnak részlegesen ellentmondó modellek korrelációs értékei magasak.**

A hasonlóságelemzési modellek esetén nincs szükség a szignifikancia-fogalomra: az egyes logikai elvárások alapvetően ellentmondás-darabszámokat mérnek (pl. stabilitás). Ennek értelmében a hasonlóságelemzési modellek számos modell-hitelességet részlegesen leíró skálán egymáshoz képest rangsorolhatók. Az ’n’-dimenziós rangsorok egy modell-értékelő hasonlóságelemzés inputjai, mely esetében az Y az azonosságot kikényszeríteni akaró konstans minden, az összes KO-feltételt meghaladó modell esetén. Az Y0-típusú hasonlóságelemzés [B20] eredménye a modellek egymáshoz képesti rangsora. A konstanst meghaladó értékek a JÓ, míg az ezt elérni nem tudó modellek a GYENGE csoportba sorolandók.

Jelen esetben az összes vizsgált modell egymással összevetve nem felel meg a KO-feltételek stabilitási, ellentmondás-mentességi szintjeinek, mert ezen tulajdonságok az eltérő modellezési keretfeltételek között lényegében teljesen szabadon változhatnak.

### WizWhy

A döntési fák logikáját követő egyik legújabb fejlesztés a döntési erőként is jellemezhető WizWhy (WW) szoftver, mely szabályok ezreit képes feltárni nagy műveletvégző sebességgel kombinálva. A vizsgálódásba azért került bele ez a megoldás is, mert a szerzők érzékeltetni szeretnék, hogy a folytonos inputok (regressziós vizsgálatok), a diszkrétté redukált, de folytonosságot szimulálni képes (hibridizáló) hasonlóságelemzés és a valóban diszkrét megoldáskeresés (WW) jelentős módszertani eltérések hordoz magában.

A WW nem képes optimum feltárására és ceteris paribus alakzatok egyszerű megjelenítésére. Tehát magára az alapkérdésre nem adhat választ. Nem képes genetikai potenciált számítani. A változók fontossága azonban itt is szimulálható. A WW képes 85 megfigyelés esetén 84 rétegben arra, hogy csoportba (küszöbérték alatti/feletti) sorolja a várható következményeket. Ez önmagában tehát a 84 rétegű bizonyítási folyamat, mely az egyes intervallumokba esés valószínűségén keresztül a hazai várható élettartamról szimulációs jelleggel tud nyilatkozni. Terjedelmi korlátok miatt jelen tanulmányba további részletek nem kerültek be…

## Vita, következtetések

### Statisztikai függvényillesztés

A statisztikai függvényillesztés viszonylagos magabiztossággal eredményt vél felmutatni tudni. Ebben az esetben a hazai prevenciós politika legitimitása látszólag nyugodt szívvel jelenthető ki.

### Hasonlóságelemzés

A hasonlóságelemzés sok-sok ellenérvet is feltárva, a statisztikai elemzésekhez hasonló részkövetkeztetésekre jutott – az alacsonyabbra pozícionálható optimum révén látszólag még masszívabbá téve a prevenciós stratégia szükségszerűségét.

A sok pro és kontra érv alapján azonban szakértői vélemény, bírói döntés nem születne, további bizonyítási kísérletek elrendelése történne meg. Ezen bizonyítási eljárások egyrészt a még kombinatorikailag lehetséges modell-alternatívák kiértékelését jelentenék, másrészt több objektum és attribútum feltárását várnák el.

### Eltérő modellezési filozófiák kölcsönhatásai

Bár nem történt jelenleg klasszikus hibridizáció: vagyis eltérő modellek részmegoldásainak keverése, de a hasonlóságelemzés már jelenleg is hibrid karakterisztikákat mutat fel (diszkrét-folytonos átmenet szimulálása logikai és numerikus műveletek kombinálásán és regressziós komponensek bevonásán keresztül). Az eltérő modellfilozófiák keveredésének színtere jelen tanulmányban az értékelési szempontrendszer egységesítése volt. Terjedelmi korlátok miatt ennek teljeskörű érzékeltetésére ugyan nem volt mód, mégis a Szerzők feltételezése szerint érzékletessé vált, hogy alapvető problémák uralkodnak a legjobb modell kiválasztása körül, melynek legfőbb oka a kényszeres túltanulás veszélye a hiba-minimalizálás kapcsán. A legjobb modell tehát nem az, mely bármilyen szempontból is a legtöbb és legjobb koefficienseket mutatja fel, mert ezen „hormonkezelt” modell-szörnyek ugyanilyen szélsőséges ellentmondásokat is rejtenek magukban.

A jelenlegi céltalansági alaphelyzetben kiutat jelenthetne, ha objektíven számon kérhető becslőmodellek esetében ezek objektív jövőbeli helyessége és a modellek bármilyen leíró karakterisztikája betárolásra kerülne tartalmi aspektusokkal és tartalom nélkül. Ezen tanulási minta alapján a leíró koefficiensek legjobb jövőbeli beválást feltáró kombinációi lennének kereshetők. Addig is, míg az első ilyen nagy volumenű vizsgálatok el nem készülnek ajánlatos, hogy a potenciális elhamarkodottságra sarkalló részmegoldásokat további pro és kontra érvekkel egészítsünk ki. Ezek alapján ugyan egyre nő azon élethelyzetek száma, melyekben nem illik tudományosan megalapozott modellekről beszélni, ellenben a politikai döntéshozatal operatív igényei megalapozottabban elégíthetők ki, mint a bértollnoki alternatívák esetében.

## Összefoglalás, ajánlások

A tanulmány részletesen, bár a kombinatorikai lehetőségekhez képest mégis csak töredékesen igyekezett bemutatni: miként is képzelhető el a modellek felett megrendezendő ún. statáriális bírósági tárgyalás. Milyen szempontok miként hatnak és milyen kölcsönhatások léteznek ezek között. A hasonlóságelemzés eszköztárának megvillantásán keresztül fény derült arra, hogy a modellek értékelésének, a konzisztenciára törekvő modell-értékelésének rétegei lényegében korlátlanok, vagyis létezik egy új (kombinatorikai, ill. hibridizált) világ az eddig közismert matematikai-statisztika mellett. A két világ között pedig számos ponton hidak verhetők, melyek a lehetőségek együttes mérlegelése elé nyitnak utakat. A hasonlóságelemzési és statisztikai modellértékelési szempontok (a KO\_feltételek kielégítése után) egy Y0-modellel [B20] vonhatók össze egyetlen egy modellrangsoroló skálává a versengő modellek tekintetében. Szimulátor-építésre és genetikai potenciálszámításra csak egy győztesnek kikiáltott modell alapján adódik lehetőség. Az egyedi modellek értékelése önmagában is anakronisztikus, hiszen a modellek egymással való összevetésére is megfogalmazhatók logikai elvárások. A modellek értékelése eddig sem volt deklaratív (pl. plot-ok kiértékelése). Az ötletadó (intuíció-generáló) sok dimenziós modellértékelést a modellideálhoz mért távolságot kifejező Y0-modell váltja fel.

A tanulmány alapján óvatosan, de vélelmezhető, hogy a hazai prevenciós politika a só-bevitel redukcióját illetően hiteles, bár további vizsgálatok elvégzése szükségesnek tűnik. Ehhez a nemzetközi szinten konszolidált adatgyűjtés mellett az ellenőrző/értelmező rétegek számos lehetőségét lehet és kell még bevonni a további elemzések körébe.

## Irodalomjegyzék

[B1]: [http://miau.gau.hu/miau2009/index\_tki.php3?\_filterText2=\*2010-01-24%2015:15](http://miau.gau.hu/miau2009/index_tki.php3?_filterText2=*2010-01-24%2015:15),

Letöltve: 2011.06.03.

[B2]: <http://miau.gau.hu/miau/147/salt.xls>,

Letöltve: 2011.06.03.

[B3]: <http://my-x.hu>, ill. <http://miau.gau.hu/myx-free>

Letöltve: 2011.06.03.

[B4]: <http://miau.gau.hu/miau/150/la150.docx>

Letöltve: 2011.06.03.

[B5]: <http://miau.gau.hu/mediawiki/index.php/Ceteris_paribus>

Letöltve: 2011.06.03.

[B6]: <http://www.google.hu/search?q=c%C3%A9ltalans%C3%A1g+t%C3%A9tele+site:miau.gau.hu&hl=en&client=firefox&rls=com.yahoo:en-US:official&prmd=ivns&ei=CILfTfPVEo-Sswajk_HPBQ&start=10&sa=N&biw=1280&bih=687>

Letöltve: 2011.06.03.

[B7]: <http://www.oeti.hu/download/motesz_stopso.pdf>,

Letöltve: 2011.06.03.

[B8]: <http://www.stopso.eu/download/programfuzet.pdf>

Letöltve: 2011.06.03.

[B9]: <http://egeszseghazugsagso.freewb.hu/so-mitosz>

Letöltve: 2011.06.03.

[B10]: <http://lb.landw.uni-halle.de/publikationen/pfupf/pfupf.htm>

Letöltve: 2011.06.03.

[B11]: <http://hu.wikipedia.org/wiki/Biomatematika> = <http://hu.wikipedia.org/wiki/Biometria>

Letöltve: 2011.06.03.

[B12]: <http://miau.gau.hu/miau2009/index.php3?x=e0&string=biom>

Letöltve: 2011.06.03.

[B13]: <http://miau.gau.hu/miau/82/kjm_hu_ecology_full.doc>

Letöltve: 2011.06.03.

[B14]: <http://miau.gau.hu/miau2009/index.php3?x=e080>

Letöltve: 2011.06.03.

[B15]: <http://miau.gau.hu/miau/31/otkastudy2.doc>

Letöltve: 2011.06.03.

[B16]: <http://miau.gau.hu/miau/116/szigma_plrf.doc>

Letöltve: 2011.06.03.

[B17]: <http://science-in-farming.library4farming.org/Food-Nutrition-Allowances/NUTRIENTS/Sodium-Potassium-and-Magnesium.html>

Letöltve: 2011.06.03.

[B18]: <http://www.oeti.hu/index.php?m1id=10&m2id=228>

Letöltve: 2011.06.03.

[B19]: <http://miau.gau.hu/miau2009/index.php3?x=e0&string=india>

Letöltve: 2011.06.03.

[B20]: <http://miau.gau.hu/myx-free/index.php3?x=e091>

Letöltve: 2011.06.03.

[B21]: <http://index.hu/tudomany/2011/05/31/who_rakkeltok_lehetnek_a_mobilok/>

Downloaded: 06.06.2011

[STAT1]: Ulrike Grömping (2006). Relative Importance for Linear Regression

in R: The Package relaimpo. Journal of Statistical Software, 17(1), 1—27

[STAT2]: R Development Core Team (2010). R: A language and environment for statistical

computing. R Foundation for Statistical Computing, Vienna, Austria.

ISBN 3-900051-07-0, URL <http://www.R-project.org/>

[STAT3]: Samprit Chatterjee és Ali S. Hadi (2006). Regression analysis by example*.*

John Wiley and Sons

[STAT4]: Schmidhuber, J. (2007). The EU Diet – Evolution, Evaluation and Impacts of the

CAP. Montreal, URL <http://www.fao.org/fileadmin/templates/esa/Global_persepctives/Presentations/Montreal-JS.pdf>

[STAT5]: Simopoulos A. (2008) The importance of the omega-6/omega-3 fatty acid ratio

in cardiovascular disease and other chronic diseases.

Experimental Biology and Medicine. Published online 11 April 2008. DOI:10.3181/0711-MR-311