

Új statisztikai módszer: kopula modellezés az egészség-gazdaságtanban

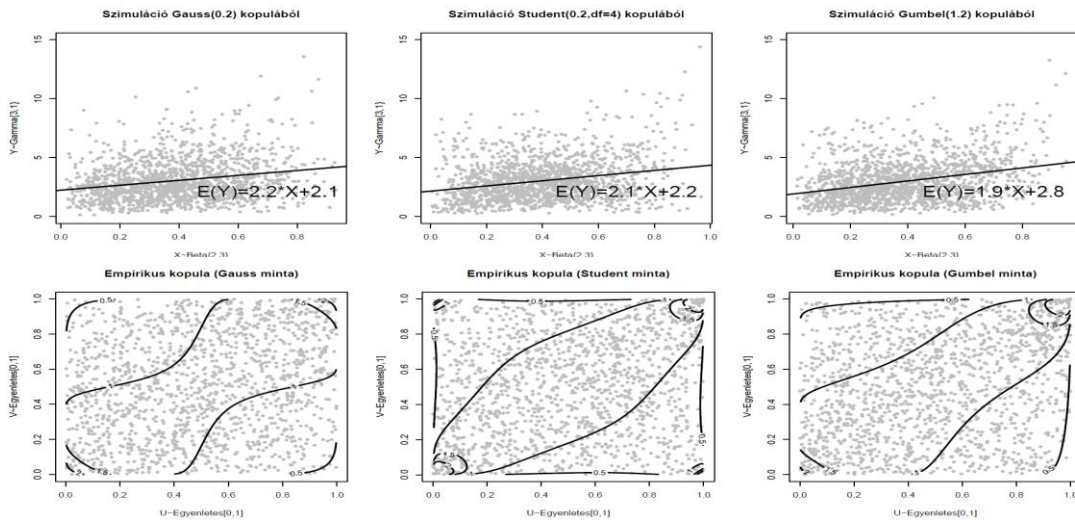
Rakonczai Pál, Balázs Tamás, Németh Tibor, Dr. Bacskai Miklós
Healthware Tanácsadó Kft.

Háttér

Napjainkban egyre elterjedtebb a többdimenziós eloszlások kopula módszerrel történő elemzése. A megnövekedett mennyiségű rendelkezésre álló információ lehetőséget nyújt arra, hogy nem csak az egyes folyamatokat, hanem a folyamatok közötti összefüggési struktúrákat is modellezni lehessen [3].

Célkitűzés:

A poszter célja, hogy megismertessen egy egyre nagyobb teret hódító modellezési eljárást a többdimenziós eloszlásokkal kapcsolatban, amely módszer segítségével elemezhetővé válik számos egészség-gazdaságtani mutató között fennálló asszociáció elemzése.



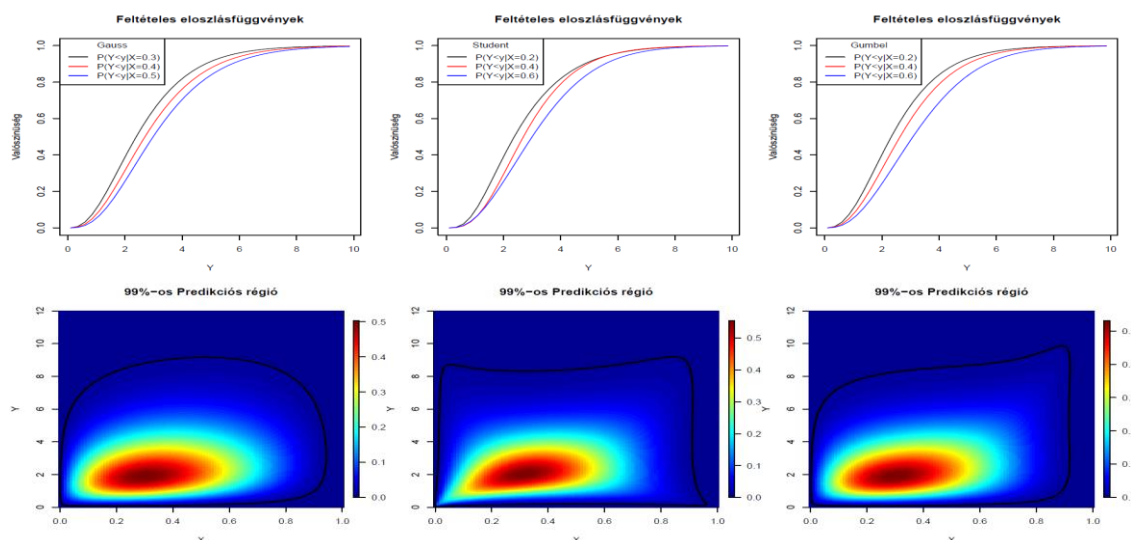
1. ábra Szimulált minták azonos béta és gamma eloszlású marginálisokkal (első sor) és a belőlük képzett pseudo-minta sűrűségfüggvény szintvonalakkal (második sor) Gauss, Student és Gumbel kopula esetén

Módszerek:

Az egészség-gazdaságtani elemzések során gyakori, hogy a vizsgálni kívánt mutatók között nem csak egyszerű lineáris kapcsolat van. Könnyen elképzelhető nemlineáris kapcsolat is, ha például magasabb értékek erősebben függenek össze, mint az alacsony értékek. Ajánlott olyan elemzési módszert választani, amely segíti a valódi kapcsolat feltárását és megfelelő információt nyújt számunkra a vizsgálni kívánt változók összefüggéséről. A változók közötti összefüggést, azaz a két változó kapcsolatát kopula-módszerekkel vizsgáljuk.

A modellek segítségével azonosíthatjuk az események, vagy változók szélsőséges, és esetleg szokatlan viselkedését is. Ez a matematikai eljárás átfogó kapcsolatmérés ad lehetőséget. Segítségükkel lehetőségünk van többdimenziós adatsorok összefüggési struktúrájának átlátható és skálafüggetlen vizsgálatára [1]. A statisztikai számítások során nincs szükség az adatsorok eloszlására vonatkozó előzetes feltételezésre, ami egyéb statisztikai eljárás során (pl.: általánosított lineáris modell) elsődleges feladat. Az elemzések során különféle eltérő tulajdonságú paraméteres kopula családokat vizsgálhatunk. A paraméteres kopulák két legnépszerűbb osztálya az elliptikus és az arkhimédészi kopulák. A különböző családok paraméter tartományai, és hogy a korrelálatlan esetnek a paraméter tartomány melyik szélé felel meg családoként eltérő.

A többdimenziós normális eloszlással definiálható normális v. Gauss kopula, és az adott szabadságfok feltételezése mellett Student-t eloszlással meghatározott t-kopula az elliptikus kopulákhoz tartoznak. A gyakorlatban másik népszerű kopulacsalád az arkhimédészi kopulák előnyös tulajdonsága, hogy a generátorfüggvényük segítségével megkonstruálhatóak, és az összefüggési struktúrák egy széles skáláját állítják elő. Néhány népszerű osztály ezen belül a Gumbel, Joe, Ali-Mikhail-Haq, Frank és Clayton. Az elemzés során a Gauss kopulát, a 4-es szabadságfokkal rendelkező t-kopulát és a Gumbel kopulát használtuk. A paraméterek illesztését pseudo-maximum likelihood eljárással végeztük. A likelihood függvényhez az ún. pseudo-mintát az empirikus eloszlásfüggvényekkel történő marginális transzformáció segítségével kapjuk, azaz az eredeti eloszlás helyett (1. ábra első sor) már az egységnyezetbe transzformált eloszlásból vett mintára végezzük az illesztést (1. ábra második sor). Ennek a vizsgálatára, hogy melyik kopulacsalád a legmegfelelőbb az összefüggés modellezésére illeszkedésvizsgálati eljárásokkal vizsgálhatók. Az illeszkedés alapján gyakran tekintik az illesztett modellnek az empirikus mintától vett távolságát, amelyet K függvényként szoktak jelölni. A választás során törekedünk a legjobban illeszkedő, azaz a legkisebb távolság mértékkel rendelkező modellt megkeresni [1,4].



2. ábra Példák: Feltételes eloszlásfüggvények (első sor) és predikciós régiók (második sor) Gauss, Student és Gumbel kopulák esetén

Eredmények és következtetés:

Néhány gyakran alkalmazott struktúrából való szimulációt követően [2] elvégezzük a paraméterek becslését. Ezt követően a lineáris regressziós módszertől eltérően (1. ábra, első sor) nem csupán a várható érték trendvonalát határozzuk meg, hanem a teljes kétdimenziós eloszlást, amelyből bármely egyszemélyes feltételes eloszlás kiszámítható (2. ábra, első sor). Különféle kopulacsaládok látványosan eltérő feltételes eloszlásokhoz vezetnek, még azonos marginális eloszlások mellett is. A 2-dimenziós konfidencia intervallumalternatívájaként meghatározhatunk adott valószínűségi (2. ábra, második sor) predikciós régiókat. A sűrűségfüggvény szintvonalain alapuló régiók megmutatják, hogy mely kompakt régiók belül esnek megfigyeléseink 99%-os valószínűséggel. A módszer hasznos segítséget ad a folyamatindikátorok és köztes eredményességi végpontok közötti kapcsolatok tisztázásában, például megállapítható, hogy az extrém nagy mérés párok között milyen

mértékű az összefüggés. A 2. ábra, második sorában balról jobbra haladva a függetlenségtől a közepes összefüggésen keresztül az erős összefüggésig mindenre láthatunk példát.

Hivatkozások:

- [1] Genest, C. and Favre, A.-C. (2007). Everything you always wanted to know about copula modelling but were afraid to ask. Journal of Hydrologic Engineering 12, 347-368.
- [2] Kojadinovic, I. and Yan, J. (2014) Package 'copula'. For the manual of the R software package see <http://cran.r-project.org/web/packages/copula/copula.pdf>
- [3] Nelsen, R. B. (2006) An introduction to Copulas. Springer, New York.
- [4] Rakonczai, P and Zempléni, A. (2008) Copulas and goodness of fit tests, Recent Advances in Stochastic Modeling and Data Analysis, p. 198-205, Word Scientific