

Kutatásmódszertani féléorák

2022. május 12.

Összefoglaló

A Kutatásmódszertani féléorák soron következő rendezvénye a statisztikai modellezés tipikus hibáival, félelmeivel és kényszereivel foglalkozott. Az előadó, Szilágyi Roland, a Gazdaságelméleti és Módszertani Intézet egyetemi docense, felvezető gondolataival megalapozta az élénk beszélgetést a résztvevők között.

Szóba került a mintavétel szerepe az eredmények alakulásában (reprezentativitás, a facebook, mint adatgyűjtési felület következményei). Az előadó a résztvevők figyelemébe ajánlotta a MTA SJTB STAB a reprezentatív minta kifejezés használatáról és a mintavételből származó adatok jellemzéséről közzétett ajánlását

(https://www.ksh.hu/docs/bemutakozas/mta/mta_stab/mta_sjtb_stab_ajanlas_reprezentativ_minta.pdf).

Felhívta a figyelmet arra, hogy a sztochasztikus kapcsolatok mindig annak a keretrendszernek a függvényei, amelyben a vizsgálat zajlik, így annak jelentőségére is, hogy mely feltételek fennállása mellett lehet működőképes a modell. Kutatói hiányosságként értékelte, ha a munka során nem indokoljuk kellőképpen, hogy miért éppen az alkalmazott változók kerülnek be a modellünkbe, és azt is aggályosnak érezte, ha a hibátag elemzésére nem térünk ki. A regressziós modellek hibátagjának vizsgálata azért is javasolt, mert a nagyobb reziduum általában arra utal, hogy legalább egy, a függő változó értékeinek alakulása szempontjából fontos indikátort kihagytunk a vizsgálatból.

Arra vonatkozóan, hogy mi legyen a kivételekkel, nem adható egyértelmű válasz, figyelembe vételük az eredményeket jelentősen torzíthatja, ugyanakkor létezésük ténye a vizsgálat szempontjából fontos információ, ezért kihagyásukat meg kell fontolni.

A beszélgetés későbbi részében az adatminőség egyes kérdéseit is megvitattuk, így például a mintanagyság hatását a szignifikanciaszint alakulására. 1000 elem fölötti mintáknál gyakori, hogy a viszonylag csekély eltérések is szignifikánsként jelennek meg a próbában, ezért erre az értékelés során figyelni kell. Statisztikus kollégáink javaslata szerint további statisztikák megadásával (átlag, box plot diagram, konfidencia intervallumok stb.) az összefüggések meglehetősen megerősíthető, vagy elutasítható. Továbbá a skálázás és az adatredukciós eljárások torzításai, adatvesztései is felmerültek.